

	Best Local Similarity	10.5%; Pred. No. 5.6e-06;			
	Matches	50; Conservative	220; Mismatches	208; Indels	0; Gaps
QY	436	ACAATGTCGATTGATAAATGCTGTTCAGAAACACACTTAATTAAGAGAAAATCATAGT	495		
DB	1446	AGAATTGGTACRRR	1387		
QY	496	TGGGATGATCAATGCAGTGCTGTGAAGTCGGAGAGAACAAATGTGGAAACCTCGCTGTG	555		
DB	1386	RRR	1327		
QY	556	CTGCCAGATGAGAAGAAATGTTCTTGTTGATCCGTCATCATACTGATATAAAGAKA	615		
DB	1326	RRR	1267		
QY	616	TTACAAAACGATTTCACAGGATTGTAATAATTATAATAGTCAATCCCTTATGGATGCTTT	675		
DB	1266	RRR	1207		
QY	676	AGCTGTTCACTGGGATATGAACAACAGACAACTGATCAATTTAGTTTAGTATAAATGAG	735		
DB	1206	RRR	1147		
QY	736	TCCACTGAAAAAGATATGAATTACAGAAACAAATGSATCCATCGAATAGACCGAACA	795		
DB	1146	RRR	1087		
QY	796	GAGGGAGATCTGTAACCATCTGTCTTCACTCTGATAGTCTACCGAGTGTCTGT	855		
DB	1086	RRRRRRRRRRRRRRRRRRRTGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGAATAA	1027		
QY	856	TCCCCTTCACAATTAAAGGATGACGGAAGTATAGGTAGAGACCCCTCCATGCTCGGA	913		
DB	1026	TTCTGTGAGCGTATGGCAAACGAAAGAAAATAGTTATAGTACCGCACTCGATGGGA	969		

RESULT 2
 US-09-543-681A-615
 ; Sequence 615, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 615
 ; LENGTH: 1575
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-615

```

RESULT 3
US-09-543-681A-707/c
; Sequence 707, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 707
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-707

Query Match 0.9%; Score 45.8; DB 4; Length 429;
Best Local Similarity 55.3%; Pred No. 0, 0019;
Matches 89; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 603 TGATAAAGGACATTACAAAAGATTTCACAGATTGTAATTAATTATATAGTCAATCCCT 662
Db 162 TCGATAATGTTAACTTAGAAAAATGATAATGTTAACTTAGAAAAATGATAATGTTAACTTAGA 103
QY 663 TATGATGCTTTTTCGTGTTCTACGTAATGAAACACAGACAAACTGATCAATTTAGTTT 722
Db 102 AAATGATAATGTAACTTAGAAAATGATAATTAATTAATGATAATGATAATGCTAACTT 43
QY 723 TAGTATAATGAGTCCACTGAAAAAGATATGAATTCAGAGA 763
Db 42 AAATATAAAAAAGGCACCTAATAAAAAAGATATAAATAGAAATA 2

```

```

RESULT 4
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

```

Query Match	0.9%; Score 44.8; DB 4; Length 832;
Best Local Similarity	10.1%; Pred. No. 0.0064;
Matches	34; Conservative 170; Mismatches 132; Indels 2; Gaps 1;
Qy	3315 ATGGGAAACTCCTGGCGTAAAGATTTCCTATCCGTCGTAGTTTGACACTTGCAGCTCA 3374 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : 360 WTWKTKTWYTTYTWTRMMWKKRWRYYWVKSTVACASRYKYTWGWWWYTWKRMMS- 302
Dd	3375 ATATCGACITTTATCCATGCCACTATTTCAGTGTCAAGATTTCGGAAGCCACTGTTTGGAGA 3434 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : 301 -TWIYCWCCKMYRGRCARWTMARGRWISYAWGKWKSMSXCTFMYKKGSTYW 243
Dd	3435 GACGGGCGCATCCATCATGAATCTTCTTGCAGACITTCAGAAAATACCAGATATACCTTGGC 3494


```
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 745:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...2515
;
US-09-221-017B-745
Query Match 0.8%; Score 41; DB 4; Length 2515;
Best Local Similarity 54.2%; Pred. No. 0.21;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 398 CTAATGGACAGACTGTAATCTAATCCAGAGATGCCACAACTGCGATTGATGAAATG 457
Db 2056 CTACCCGTCGAGAGCTCTCTAATAAACAAGAGATGAAAAAACAAGTAATTAATGAATAC 1997
QY 458 CTGTTGCAGAGACAGAGTAATTAAGAGAACTAGTTGGGATGATCAATGCAAGTCTG 517
Db 1996 ATTATGGAATTATCGAGTGTCTTAAAGAGAAACAATTTTAAGGTAGTTGAATTTCAAGATA 1937
QY 518 TTGAAGTGGAGAGAGAAATGTGAACCTGG 550
Db 1936 TTGAAGATGCAAGAGATAAGTGTGAACATCTGG 1904

RESULT 8
US-08-257-073-2
; Sequence 2, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-257-073-2
Query Match 0.8%; Score 39.6; DB 1; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGGTGATAGTGGAGACAGTGTGTTGGATTGGCAGATCGAGGTCTAGATTAAAGGAA 1486
Db 1568 ATAAAGATAATAAAGGAAAGGTGTCGATACATGTTGATACAACTTAGAAAAAGAGATA 1627
QY 1487 CTTGCATTAGTCAAGTGAAGATGCTGATTTCTCCACTGTTTATACACACACCAGCAGCAA 1546
Db 1628 CTTTATCATATGATCACTACGATAATATGTTTGTAAATAAGAAATTTGTACACATTAA 1687
QY 1547 ATATCTATCTAATGGTGTGATCTCTATGGAATGCAAGACCAGGTGTTCTTTTGTTC 1606
Db 1688 AAGATGAAATAATTTGTATATCTAATCTTCAAGTTGAAGATCAAGGTAAATTTGTATACTT 1747
QY 1607 CAAAGACTTTACCTCCAAAGAGATTCAGTAAC 1640
Db 1748 CATGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 9
US-08-184-009-119
; Sequence 119, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
```



```
;
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-184-009-119

Query Match      0.8%; Score 39.6; DB 2; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGGTGATAGTGGAGGACAGTGTGTTGGATTGGCAGATGCGAGGTCTAGATTAAAGGAA 1486
Db 1568 ATAAAGATAATAAGGAAAGGTGTCGTACATGTTGATACACACTTTAGAAAAGAGATA 1627
QY 1487 CTTGCATTAGTGAAGTGAAGAAATGTTGATTTCTCCACTGTTATAGACACACCGAGCAA 1546
Db 1628 CTTTATCATATGATACTCAGATAATATGTTTGTATAAAGAAATATTGTAACAGATTAA 1687
QY 1547 ATTATCTATCTAATGTTGATTCCTATGGAATCCAGACCCAGGTGTTCTTTTGTTC 1606
Db 1688 AAGATGAAATAATATTGTATATCTAATCTTCAAGTTGAAGATCAAGGTAATTGTGATACTT 1747
QY 1607 CAAGACTTTACCTCCAAAGAAAGATTCAAGTAAC 1640
Db 1748 CATGGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 10
US-08-458-356-119
; Sequence 119, Application US/08458356
; Patent No. 5942235
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.356
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-184-009-119

;
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-458-356-119

Query Match      0.8%; Score 39.6; DB 2; Length 2981;
Best Local Similarity 49.1%; Pred. No. 0.65;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1427 ATGGTGATAGTGGAGGACAGTGTGTTGGATTGGCAGATGCGAGGTCTAGATTAAAGGAA 1486
Db 1568 ATAAAGATAATAAGGAAAGGTGTCGTACATGTTGATACACACTTTAGAAAAGAGATA 1627
QY 1487 CTTGCATTAGTGAAGTGAAGAAATGTTGATTTCTCCACTGTTATAGACACACCGAGCAA 1546
Db 1628 CTTTATCATATGATACTCAGATAATATGTTTGTATAAAGAAATATTGTAACAGATTAA 1687
QY 1547 ATTATCTATCTAATGTTGATTCCTATGGAATCCAGACCCAGGTGTTCTTTTGTTC 1606
Db 1688 AAGATGAAATAATATTGTATATCTAATCTTCAAGTTGAAGATCAAGGTAATTGTGATACTT 1747
QY 1607 CAAGACTTTACCTCCAAAGAAAGATTCAAGTAAC 1640
Db 1748 CATGGATTTTCTTCAAAATATCATTTAGAAAC 1781

RESULT 11
US-08-460-736-119
; Sequence 119, Application US/08460736
; Patent No. 6265189
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,736
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/184,009
; FILING DATE: 19-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2981 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-460-736-119
```


This Page Blank (uspto)

Result No.	Score	Query Match	Length	DB	ID	Description
1	1830	37.8	1830	9	US-09-764-878-355	Sequence 355, App
2	1830	37.8	1830	9	US-09-764-878-356	Sequence 356, App
3	1830	37.8	1830	15	US-10-079-854-355	Sequence 355, App
4	1830	37.8	1830	15	US-10-079-854-356	Sequence 356, App
5	620	12.8	703	15	US-10-066-543-681	Sequence 681, App
6	570.4	11.8	610	9	US-09-764-878-75	Sequence 75, Appl
7	570.4	11.8	610	15	US-10-079-854-75	Sequence 75, Appl
8	559	11.6	5402	15	US-10-079-977-11	Sequence 11, App
9	432	8.9	501	15	US-10-066-543-1215	Sequence 1215, App
10	401	8.3	401	15	US-10-066-543-1369	Sequence 1369, App
11	260.2	5.4	498	10	US-09-918-995-9986	Sequence 9986, App
12	220.8	4.6	969	16	US-10-264-237-394	Sequence 394, App
13	203.4	4.2	475	10	US-09-918-995-3724	Sequence 3724, App
14	154	3.2	437	9	US-09-983-965-4329	Sequence 4329, App

Qy	381	GGAGGAGTCACTGCTTAATGACAGACTGTAACTTAAATCCAGAGATTGCCACAAT	440
Db	181	GGAGGAGTCACTGCTTAATGACAGACTGTAACTTAAATCCAGAGATTGCCACAAT	240
Qy	441	GTGGATTGATGAAAATGCTGTTGCGAGAGACCAAGTTAAATTAAGAGAAAATATAGTTGGGA	500
Db	241	GTGGATTGATGAAAATGCTGTTGCGAGAGACCAAGTTAAATTAAGAGAAAATATAGTTGGGA	300
Qy	501	TGATCAATGCAGTGCCTGTTGAAGTGGGAGAGAAAATGTGAAAACCTCGCTGTGTCGC	560
Db	301	TGATCAATGCAGTGCCTGTTGAAGTGGGAGAGAAAATGTGAAAACCTCGCTGTGTCGC	360
Qy	561	AGATGAGAAATGTTCTGTTGTCAGCGCTGATGCAATTAAGTGTATAAAGACACATTACA	620
Db	361	AGATGAGAAATGTTCTGTTGTCAGCGCTGATGCAATTAAGTGTATAAAGACACATTACA	420
Qy	621	AAACGATTTACAGGATTGTAATAATTAATTAAGTCAATCCCTTATGGATGCTTTTAGCTG	680
Db	421	AAACGATTTACAGGATTGTAATAATTAATTAAGTCAATCCCTTATGGATGCTTTTAGCTG	480
Qy	681	TTCACTGGATTAATGAAAACAGACAAACTGATCAATTTAGTTTTAGTATATAAAGAGTCCAC	740
Db	481	TTCACTGGATTAATGAAAACAGACAAACTGATCAATTTAGTTTTAGTATATAAAGAGTCCAC	540
Qy	741	TGAAAAAGATATGAATTTACAGAGAAAACAAATGGATCCATTGAAATAGACCGAAAAACAGGG	800
Db	541	TGAAAAAGATATGAATTTACAGAGAAAACAAATGGATCCATTGAAATAGACCGAAAAACAGGG	600
Qy	801	GAGATCTGTTAAACCATCTGTGTCCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC	860
Db	601	GAGATCTGTTAAACCATCTGTGTCCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC	660
Qy	861	TTCACAATTTAAAGGATCAGCGAAGTATAGTATAGAGACCCCTCCATGTCGCGATTACAAG	920
Db	661	TTCACAATTTAAAGGATCAGCGAAGTATAGTATAGAGACCCCTCCATGTCGCGATTACAAG	720
Qy	921	TTTAAACGGTTGATTCAGTAAATCTCATCCCAGGGAACAGATGGATGTCCTGTTAAAAA	980
Db	721	TTTAAACGGTTGATTCAGTAAATCTCATCCCAGGGAACAGATGGATGTCCTGTTAAAAA	780
Qy	981	GCAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCCTAGGACAGA	1040
Db	781	GCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCCTAGGACAGA	840
Qy	1041	TCTAGGGAGTCCAAATTCCTTTTCCCACATGAGTGGGGGATTTGATGAAAAAAGAGCC	1100
Db	841	TCTAGGGAGTCCAAATTCCTTTTCCCACATGAGTGGGGGATTTGATGAAAAAAGAGCC	900
Qy	1101	AGCAGAGGAGACACCACTAGAGAAATCCCTCCGGTCTGGTTTACCTTTGCTTCTCAAAAC	1160
Db	901	AGCAGAGGAGACACCACTAGAGAAATCCCTCCGGTCTGGTTTACCTTTGCTTCTCAAAAC	960
Qy	1161	AGACATGCCTAATGGGCTGCGAAGGAATATGAATGTGAACGGTGTTCAGATTGCCCTGT	1220
Db	961	AGACATGCCTAATGGGCTGCGAAGGAATATGAATGTGAACGGTGTTCAGATTGCCCTGT	1020
Qy	1221	GCCTAATGAAGTTAGGGCTGATGAAAATGAAGGTTNTGAACATGAAGAACTCTTGGCAC	1280
Db	1021	GCCTAATGAAGTTAGGGCTGATGAAAATGAAGGTTNTGAACATGAAGAACTCTTGGCAC	1080
Qy	1281	TACAGAAATCTTAATATGACAGAGCAATTTCTCTGAATCTCAGGACATGACTAATTGGAA	1340
Db	1081	TACAGAAATCTTAATATGACAGAGCAATTTCTCTGAATCTCAGGACATGACTAATTGGAA	1140
Qy	1341	GTTGACTAAATTAATAGAGATGAATGATAGCCAAAGTAAACGAAGAAAGGAAAGTTTCT	1400
Db	1141	GTTGACTAAATTAATAGAGATGAATGATAGCCAAAGTAAACGAAGAAAGGAAAGTTTCT	1200
Qy	1401	ACAGATTAGTCAGCTGAGGACACTTAATGGTGATAGTGGAGCAGTGTGTGGATTGGC	1460
Db	1201	ACAGATTAGTCAGCTGAGGACACTTAATGGTGATAGTGGAGCAGTGTGTGGATTGGC	1260
Qy	1461	AGATGCAGGCTTAGATTTTAAAGGAACTTGCATTAGTGAAGAGAAATGATGTTCTC	1520

Db	1261	AGATGCAAGTCTAGATTATAAGGAACCTTGCATTAGTGAAGTGAAGTGAATGTGATTTCTC	1320
Qy	1521	CACCTGTTATAGACACACCAGCAGCAAAATTATCTATTAATGTTGTGATTCCTATGGAAT	1580
Db	1321	CACCTGTTATAGACACACCAGCAGCAAAATTATCTATTAATGTTGTGATTCCTATGGAAT	1380
Qy	1581	GCAAGACCCAGTGTTCCTTTCTGTTCCAAAGACTTTACCCCTCCAAAGAGATTCAGTAAAC	1640
Db	1381	GCAAGACCCAGTGTTCCTTTCTGTTCCAAAGACTTTACCCCTCCAAAGAGATTCAGTAAAC	1440
Qy	1641	AGAAGAAAAAGAAATAGAGGAAAAAGCAAGTCAGAAATGCTACTCAAATATTTATGAACAGAG	1700
Db	1441	AGAAGAAAAAGAAATAGAGGAAAAAGCAAGTCAGAAATGCTACTCAAATATTTATGAACAGAG	1500
Qy	1701	AGAAATGAGGCCACAGAGGGAGTGACTACTCTTTTAAACAGCACTGGTGACCTTAATGAA	1760
Db	1501	AGGAAATGAGGCCACAGAGGGAGTGACTACTCTTTTAAACAGCACTGGTGACCTTAATGAA	1560
Qy	1761	GAAAAATATTTACATAATTTCTGTAGTCAAGTTCATCAGTGTCTGGGCAATCTTCCCC	1820
Db	1561	GAAAAATATTTACATAATTTCTGTAGTCAAGTTCATCAGTGTCTGGGCAATCTTCCCC	1620
Qy	1821	CAAGGTAGTAGCAAGCCTGCATCTATCAGTGTTCCTTTTGGTGTCAGACCCCAAGCA	1880
Db	1621	CAGGTAGTAGCAAGCCTGCATCTATCAGTGTTCCTTTTGGTGTCAGACCCCAAGCA	1680
Qy	1881	ACCTTCTAATCTTAAACTTCAAATTCCAAAGCCATATCAGACCATTTACAAAATGACTT	1940
Db	1681	ACCTTCTAATCTTAAACTTCAAATTCCAAAGCCATATCAGACCATTTACAAAATGACTT	1740
Qy	1941	TCCTGCAAAACAGTGGAAATTAATCTAAAAATAAAAAATGATATTCCTGGGAAAACAAAAT	2000
Db	1741	TCCTGCAAAACAGTGGAAATTAATCTAAAAATAAAAAATGATATTCCTGGGAAAACAAAAT	1800
Qy	2001	AGGGAAAACTCAGCAACCAATGTATGCAG 2030	
Db	1801	AGGGAAAACTCAGCAACCAATGTATGCAG 1830	

RESULT 3

US-10-079-854-355

; Sequence 355, Application US/10079854

; Publication No. US20030054368A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PAL21C1

; CURRENT APPLICATION NUMBER: US/10/079,854

; PENDING FILING DATE: 2002-02-22

; Prior Application removed - See File Wrapper for Palm

; NUMBER OF SEQ ID NOS: 428

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 355

; LENGTH: 1830

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-079-854-355

Query Match 37.8%; Score 1830; DB 15; Length 1830;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	201	AGATGAACACAGTTTCTTCTACTTTATGGATACAAAGTGGAAATAGATTCAGATCCCCC	260
Db	1	AGATGAACACAGTTTCTTCTACTTTATGGATACAAAGTGGAAATAGATTCAGATCCCCC	60
Qy	261	TTCTCACCGGCTGTCAATTTAACCTACTTTGGCCAGTGTGAATGAATCTGCAGTTTCTAA	320
Db	61	TTCTCACCGGCTGTCAATTTAACCTACTTTGGCCAGTGTGAATGAATCTGCAGTTTCTAA	120
Qy	321	TGAGTCACAACACCAACTGAAAGTCTTCCCTGGCTCATTCAGTCCCTGACACACAGA	380

RESULT 3

```

US-10-079-854-355
; Sequence 355, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 355
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-355

```

```
Query Match          37.8%; Score 1830; DB 15; Length 1830;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1830: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

201 AGATGAACAGCTTCTTCTACTTTATTGGATACAAAGTGAATAAGATTCTAGATCCCC
261 TTCTCACCGGCTGCATTTTAAACCTACTTTGGCCAGTGAATGAATCTGCAGTTTCTAA 320
1 AGATGAACAGCTTCTTCTACTTTATTGGATACAAAGTGAATAAGATTCTAGATCCCC 60
61 TTCTCACCGGCTGCATTTTAAACCTACTTTGGCCAGTGAATGAATCTGCAGTTTCTAA 120
321 TGAGTCAACAACCAACTGAAAGCTTCTCTCCCTGGCTCATTTAGCTCCCTCTGACCAAGA 380

321 TGAGTCACACCAAACTGAAAGTCTTCTCCCTGGCTCATTCAGCTCCCTGACACAGA 380
Db
121 TGAGTCACACCAAACTGAAAGTCTTCTCCCTGGCTCATTCAGCTCCCTGACACAGA 180
Qy
381 GGAAGAGGATCACTGTGCTAAATGGAACAGACATGTAATCTAAATCCAGAGATTGCCACAAAT 440
Db
181 GGAAGAGGATCACTGTGCTAAATGGAACAGACATGTAATCTAAATCCAGAGATTGCCACAAAT 240
Qy
441 GTGGATTGATGAAATGCTGTGTCACAGACAGACAGTAAATTAAGAGAAAATATAGTTGGGA 500
Db
241 GTGGATTGATGAAATGCTGTGTCACAGACAGACAGTAAATTAAGAGAAAATATAGTTGGGA 300
Qy
501 TGATCAATGCACTGCTGTGTAAGTGGGAGAGAGAAATGTGAAACCTCGCTGTCTGCC 560
Db
301 TGATCAATGCACTGCTGTGTAAGTGGGAGAGAGAAATGTGAAACCTCGCTGTCTGCC 360
Qy
561 AGATGAGAGAAATGCTGTGTTGTTAGCGCTCATGCACTGTAAGTAAAGAGACATTACA 620
Db
361 AGATGAGAGAAATGCTGTGTTGTTAGCGCTCATGCACTGTAAGTAAAGAGACATTACA 420
Qy
621 AAACGATTTACAGGATTTGAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 680
Db
421 AAACGATTTACAGGATTTGAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 480
Qy
681 TTCACTGGATTAAGAAACAGACAACTGATCAATTTAGTTTATGATTAATAGTCCAC 740
Db
481 TTCACTGGATTAAGAAACAGACAACTGATCAATTTAGTTTATGATTAATAGTCCAC 540
Qy
741 TGAAGAAGATATGAATTCAGAGAAACAAATGGATCCATTGAATAGACCCGAAACAGAGG 800
Db
541 TGAAGAAGATATGAATTCAGAGAAACAAATGGATCCATTGAATAGACCCGAAACAGAGG 600
Qy
801 GAGATCTGTTAACCATCTGTCTCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 860
Db
601 GAGATCTGTTAACCATCTGTCTCTACTTCATCTGATAGTCTAGCCAGTGTCTGTTCCCC 660
Qy
861 TTCAACAATTAAGAGATGACGGAATATAGTATGAGACCCCTCCATGTCGCGATTACAAG 920
Db
661 TTCAACAATTAAGAGATGACGGAATATAGTATGAGACCCCTCCATGTCGCGATTACAAG 720
Qy
921 TTTAACCGTTGATTCAGTAATCTCATCCAGGAAACAGATGATGATGCTCTGTTAAAAA 980
Db
721 TTTAACCGTTGATTCAGTAATCTCATCCAGGAAACAGATGATGATGCTCTGTTAAAAA 780
Qy
981 GCAAGAGAACTATATACAGATGAGACCTCATCTGCAAAATCAGCTCTCTAGGACAGA 1040
Db
781 GCAAGAGAACTATATACAGATGAGACCTCATCTGCAAAATCAGCTCTCTAGGACAGA 840
Qy
1041 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGGATTTTGATGAAAAAGAGCC 1100
Db
841 TCTAGGAGTCCAAATTCCTTTTCCCATGAGTGGGGATTTTGATGAAAAAGAGCC 900
Qy
1101 AGCAGAGAGAGCACCCTGAAGAAATCCCTCGGTCGTGTTTACCTTTGCTCTCAACC 1160
Db
901 AGCAGAGAGAGCACCCTGAAGAAATCCCTCGGTCGTGTTTACCTTTGCTCTCAACC 960
Qy
1161 AGACATCCCTAATGGTCTCGAAGGATATGATGTAAGCGTGTTCAGATTGCTTGT 1220
Db
961 AGACATCCCTAATGGTCTCGAAGGATATGATGTAAGCGTGTTCAGATTGCTTGT 1020
Qy
1221 GCCTAATGAAGTATAGGCTGATGAAAAATGAAGTTATGAACATGAAGAAATCTTGGCAC 1280
Db
1021 GCCTAATGAAGTATAGGCTGATGAAAAATGAAGTTATGAACATGAAGAAATCTTGGCAC 1080
Qy
1281 TACAGAAATCTTAATGACAGAGATTTCTCTGAAATCTCAGGACATGATGTTGGAA 1340
Db
1081 TACAGAAATCTTAATGACAGAGATTTCTCTGAAATCTCAGGACATGATGTTGGAA 1140
Qy
1341 GTTCACATAAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db
1141 GTTCACATAAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy
1401 ACAGATTAGTCAGCCTGAGGACACTAATGTTGATGATGAGGAGCAGTGTGTTGATGGC 1460

1201 ACAGATTAGTCAGCCTGAGGACACTAATGTTGATAGTGGAGCAGTGTGTTGATGGC 1260
Qy
1461 AGATGCAAGTCTAGATTAAAGGAACTTGCAATTAGTAAAGTGAAGAAATGTTGATTTCTC 1520
Db
1261 AGATGCAAGTCTAGATTAAAGGAACTTGCAATTAGTAAAGTGAAGAAATGTTGATTTCTC 1320
Qy
1521 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGTTGTTGATTCCTATGAAAT 1580
Db
1321 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGTTGTTGATTCCTATGAAAT 1380
Qy
1581 SCAGACCCAGTGTCTTTTGTTCAGAGACATTTACCTCCCAAGAGATTCAGTAAC 1640
Db
1381 SCAGACCCAGTGTCTTTTGTTCAGAGACATTTACCTCCCAAGAGATTCAGTAAC 1440
Qy
1641 AGAAGAAAAGAAATAGAGGAAACAGTCAAGATGCTTACTCAAAATATTTATGAACAGAG 1700
Db
1441 AGAAGAAAAGAAATAGAGGAAACAGTCAAGATGCTTACTCAAAATATTTATGAACAGAG 1500
Qy
1701 AGGAAATGAGCCACAGAGGAGTGGACTACTTTTAAACAGCAGCTGTTGACCTTAATGAA 1760
Db
1501 AGGAAATGAGCCACAGAGGAGTGGACTACTTTTAAACAGCAGCTGTTGACCTTAATGAA 1560
Qy
1761 GAAAAATTTATTTACATAATTTTCTGTAGTCAAGTTCCATCAGTGTGTTGGCAATCTTCCC 1820
Db
1561 GAAAAATTTATTTACATAATTTTCTGTAGTCAAGTTCCATCAGTGTGTTGGCAATCTTCCC 1620
Qy
1821 CAAGTATGACCAAGCTGCCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 1880
Db
1621 CAAGTATGACCAAGCTGCCATCTATCAGTGTCTCTTTTGGTGGTGAAGACCCCAAGCA 1680
Qy
1881 ACCTTCTAATCTTAACTTCAAAATTCRAAGCCATTTACAGCATTATACAAATGACTT 1940
Db
1681 ACCTTCTAATCTTAACTTCAAAATTCRAAGCCATTTACAGCATTATACAAATGACTT 1740
Qy
1941 TCCTCAACAGCTGGAATAATATCTAAATAAATAAATAATGATATCTTGGAAAGCAAAAT 2000
Db
1741 TCCTCAACAGCTGGAATAATATCTAAATAAATAAATAATGATATCTTGGAAAGCAAAAT 1800
Qy
2001 AGGGAAAACTCAGCAACCAATGATGCGAG 2030
Db
1801 AGGGAAAACTCAGCAACCAATGATGCGAG 1830

RESULT 5

US-10-066-543-681
; Sequence 681, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Maxgarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 681
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

Query Match 11.8%; Score 570.4; DB 9; Length 610;

Query Match 11.8%; Score 570.4; DB 15; Length 610;
Best Local Similarity 99.2%; Pred. No. 7.5e-146;
Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2

QY 1 GCATCTGAATCACAGAGACTGGCTGTGTGTGCAGCAGACATCATCAGTAAGCACCGAGA 60
Db 1 GCATCTGAATCACAGAGACTGGCTGTGTGTGCAGCAGACATCATCAGTAAGCACCGAGA 60
QY 61 AGCTGTGTCCTTATCACTGTGTGAAGGGGAAAAAGGTTTAAACAAGTCTCTTAACTGTGTG 120

Qy	121	TTTCTCTACCGATGGAGAATTACTTCCAAGCAGAAGCTTACAACCTGGGACAAGGTGTTA	150
Db	179	TTTCTCTCACCGATGGAGAATTACTTCCAAGCAGAAGCTTACAACCT-GGACAAGGTGTTA	237
Qy	181	GATGAATTTTGAAACAAACGAAGATGAAACAGATTCTCTTCTACTTTATTGGATCAAAAGTGG	240
Db	238	GATGAATTTTGAAACAAACGAGATGAAACAGTTCTTCTACTTTATTGGATCAAAAGTGG	297
Qy	241	AATAAGATTCTAGATCCCGCTTCTCTACCGGCTGTCAATTTAAACCTACTTTGGCCAGTGTG	300
Db	298	AATAAGATTCTAGATCCCGCTTCTCTACCGGCTGTCAATTTAAACCTACTTTGGCCAGTGTG	357
Qy	301	AATGAATCTGCGAGTTCTTAATGAGTCAACAACCAACTGAAAGTCTTCTCCCTGGCTCAT	360
Db	358	AATGAATCTGCGAGTTCTTAATGAGTCAACAACCAACTGAAAGTCTTCTCCCTGGCTCAT	417
Qy	361	TCAGCTCCCTGCACACAGAGNAGGATCACTGTCTAATGGACAGAGTGTATATCTA	420
Db	418	TCAGCTCCCTGCACACAGAGNAGGATCACTGTCTAATGGACAGAGTGTATATCTA	477
Qy	421	AATCCAGAGATTGCCACAATGTGG	444
Db	478	AATCCAGAGATTGCCACAATGTGG	501

RESULT 10

US-10-066-543-1369/c

; Sequence 1369, Application US/10066543

; Publication No. US20030087818A1

; GENERAL INFORMATION:

APPLICANT: Jiang, Yuqiu

APPLICANT: PYLE, RUTH A.
APPLICANT: YU, JIANQIANG

APPLICANT: Indirias, Carol Yoseph

APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Carter, Darrick

APPLICANT: Fanger, Gary R.

APPLICANT: SMITH, CAROLE L.
APPLICANT: Durham Monksville

APPLICANT: John A. Stolk
AFFIDAVANT: Margaret Williams

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

1. TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543
 ; CURRENT APPLICATION NUMBER: US/10/066,543

; CURRENT FILING DATE: 2002-01-31
 ; NUMBER OF SEC ID NOS: 2417

```

; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 1369

```

; LENGTH: 401

```

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-066-543-1369

Query Match

Best Local Similarity 0.0%, **Pred. No.** 2.6e-99;
Best Local Similarity 100.0%; **Pred. No.** 2.6e-99;

Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

369 CCTGACCACAGAGGAAGAGGATCACTGTGCTAATGGACAGGACTGTAATCTAAATCCAGA 428

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

DB 401 C C T G A C C A C A G A G A G G A A G G A T C A C T G T G C T A A T G G A C A G G A C T G T A A T C T A A A T C C A G A 342

QV 429 GATTGCCACAATGTGGATTGATGAAAATGCTGTGCAGAAAGAC CAGTTAATTAA GAGAAA 488

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

341 GATTGCCACAATGTGGATTGATGAAATGCTGTTGCAGAAGACCAGTTAATTAAAGAGAA 282

489 CTATAGTTGGGATGATCAATGCAGTCTGTTGAAGTGGGAGAGAGAAATGTGGAACCT 548
QY

[illegible]

27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054

549 GGCTTGCTGCCAGATGAGAAGATTCTTGTGTAGCCGTCATGCATAACTGTGATAA 608

100

D_b 221 GGCTTGCTGCCAGATGAGAAGATGTTCTTGTGTAGCCGTCATGCATAACTGTGATAA 162

609	QY	AAGGACATTACAAAAAGATTACAGGATTGTAAATAATTAAATAGTCAATCCCTTATGGA	668
161	Db	AAGGACATTACAAAAAGATTACAGGATTGTAAATAATTAAATAGTCAATCCCTTATGGA	102
669	QY	TGCTTTTAGCTGTTTCACCTGGAATATGAAAAAGACACAACTGATCAATTTAGTTTTAGTAT	728
101	Db	TGCTTTTAGCTGTTTCACCTGGAATATGAAAAAGACACAACTGATCAATTTAGTTTTAGTAT	42
729	QY	AAATGAGTCCACTGAAAAAGATGTAATTCAGAGAAACAAA	769
41	Db	AAATGAGTCCACTGAAAAAGATGTAATTCAGAGAAACAAA	1

RESULT 11

```

US-09-918--995-9986
; Sequence 9986, Application US/0918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSeq for Windows Version 3.0
;   SEQ ID NO 9986
;   LENGTH: 498
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(498)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918--995-9986

```

Query Match	5.4%;	Score 260.2;	DB 10;	Length 498;	
Best Local Similarity	97.1%;	Pred. No. 1.8e-60;			
Matches 265;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0	
QY	287	CTTTGGCCAGTGTGAATGAATCTGCAGATTTCTAATGAGTCACAACCACTGAAAGTCT	346		
Db	226	CTTGGCCACGTGTGAATGAATCTGCAGTTTCTAATGAGTCACAACCACTGAAAGTCT	285		
QY	347	TTCTCCCTGGCTCATTCAGCTCCCTGACACAGAGGAGAGATCACTGTGCTAATGGAC	456		
Db	286	TTCTCCCTGGCTCATTCAGCTCCCTGACACAGAGGAGAGATCACTGTGCTAATGGAC	345		
QY	407	AGGACTGTAACTTAAATCCAGAGATTGCCAATGTGGATGTGATAAATGCTGTTGCAG	466		
Db	346	AGGACTGTAACTTAAATCCAGAGATTGCCAATGTGGATGTGATAAATGCTGTTGCAG	405		
QY	467	AAGACCAGTTAAATTAAAGAAAACTATAGTTGGGATGATCAATGCAGTGTCTGTGAAATGG	526		
Db	406	AAGACCAGTTAATTAGAGAACTATAGTTGGATGATCAATGCAGTGTCTGTGAAATGG	465		
QY	527	GAGAGAAGAAATGTGGAAACCTGGCTTGTCTGC	559		
Db	466	GAGAGAAGAAATGTGGAAACCTGGCTTGTCTGC	498		

RESULT 12

US-10-264-237-394
; Sequence 394, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237

```

; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 394
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (845)..(845)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871)..(871)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (897)..(897)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (899)..(899)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-264-237-394

Query Match 4.6%; Score 220.8; DB 16; Length 969;
Best Local Similarity 59.4%; Pred. No. 2.1e-49;
Matches 442; Conservative 1; Mismatches 292; Indels 9; Gaps 4;

2824 QY CTTATTCCTGAAGTGGCTTCCTCCCAATCTCATCTCCACTGGTGTAAGAAG--AGAC 2880
2825 DB CTTAATGATGAGGACAGTTGCGCCCACTCTCTGGTGTGATCTGGAGAAAAGGATCAGTG 261
2826 QY TATGCTGTGGAGAGAAAACCATCAGATTTTCAGTAATGCAGCAGTTGGAGGATGCTGGC 2940
2827 DB CCTGTAGTAGAAGACATCCATCTCATGAGCAGATCAATTTGCTTCTTCAAGGTGAAGGC 321
2828 QY COTGACCCACTGTGATTTGTTTTAAATGCAATTTGTGTCTCAATGGTTAAATTTGAAAT 3000
2829 DB TTTTCATCTGTTACATTTTGTCTAAATGCTAAATCTACTCGTAATGTCAAATTCATATT 381
3001 QY TATGTGAACAGGAAGTGTGTGGTGTTCACAAACAAAGGGAATGCATGCTGGGTCAAGTCT 3060
3002 DB TAT-TCTTCAGACAAATATGTTACTTTTCAACCAATGGATTCGAGCTGGACAGG--- 437
3061 QY GAGATPAGTCATTCCTTCTACAGTGTGTTACGGATGAAAAGTGTGTGCGCAAGGATATCTTT 3120
3062 DB CAGAATTATATCTATTTGTTATGTTTGTCCAAATGAAGATACTATTCTTAAGGACATCTTC 497
3121 QY AATCACTTTGTGCAGCTTTATCGGATGCTCTGGCAGGGAATGTGGTGAGCAACTTTGGGA 3180
3122 DB AGACTATTTATCACCATATPAAGATGCTCTTAAAGGAAAATACATAGAAAACCTTGGAC 557
3181 QY CATTCCTCTTTCAGTCAAAGTTTCCTTGGCAGTAAAGAACATGGTGGATCTTTATATGTG 3240
3182 DB AATATTACCTTTACTGAGAGTTTCTCAGTAGCAAGGATCAGCGAGGATTCCTGTGTTATT 617
3241 QY ACATCTACTACAGTCTCAGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGTTGGG 3300
3242 DB ACACCTACTTTTCAGAACTTGATGATCTCTCATTTACCAAGTAATCCTTTCTTTGTGGA 677
3301 QY ATTCTTATCCAGAAATGGAAACTCCTTGGGTGTAAGTATTTTCCATATCCGTCTGATGTTG 3360
3302 DB ATTCTTATCCAGAACTTGAGATTCCTGGGCAAAAGGTTTTTCCCTATGGGTTTAATGTTG 737
3361 QY AGACTTGGAGCTGAATATCGCTTTTATCCATGCCCACTATTTCAGTGTCAAGATTCGGAAG 3420
3362 DB AGATTTGGGTCAGATATTAAGCATATCTGTCCTCTTAACAAAGATCAGAACGGCAAAA 797

```

QY 3421 CCATGTTTGGAGACGGGCATACATCATGTAATCTCTTGACAGCTTCAGAAATAC 3480
Db |||||
798 CTTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACCTTGTGACTTCGAAATAC 857
QY 3481 CAGTATACCTCCAGTAGTTCAGGTTTGGTGG--TTGATATGAAGTTCGGAACCTA 3538
Db |||||
858 CAGTATACCTTCGTTTAAATAAATCAACTGTGAATCCNTTGAATGGGAAAAAACC 917
QY 3539 GCATCAAAATCCCAAGCAACAGAT 3562
Db 918 GCCTTAAATCCCGGAAAGTT 941

RESULT 13
US-09-918-995-3724
; Sequence 3724, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3724
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3724

Query Match 4.2%; Score 203.4; DB 10; Length 475;
Best Local Similarity 64.6%; Pred. No. 7.9e-45;
Matches 303; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 3025 TTCACACCAAGGAATGCTAGTCAGTGGTCTGAGATGATCATCTTCTACAGTGT 3084
Db 3 TTTCAACCAATGGATTCATGCTGGTACGAGCAGCAGAAATATATCTATTGTTAATG 62
QY 3085 TTACGGATGAAAAGTGTTCCTCCAAAGGATATCTTTAATCATCTTGTGCAGCTTTATCGG 3144
Db 63 TTGCCAATGAAGTACTATTCTTAAGGACATCTTCAGACTATTATCACCATATATAG 122
QY 3145 GATGCTTGGCAGGAATGCTGAGCAACTTGGGACATTCCTTCTTCAGTCAAGTTTC 3204
Db 123 GATGCTCTAAAGGAAAAATACATAGAAAACCTTGGACAAATATTACCTTTTACTGAGAGTTT 182
QY 3205 CTTCGCTAAGAACATGTTGGATCTTATATGTGACATCTACCTACAGTCACTGCAA 3264
Db 183 CTGATAGCAGGATACGAGGAAATCTGTTTATACACTACTTTTCAGAACTTGAT 242
QY 3265 GACCTAGTACTCCAAACCCACCTTACTTGTGGGATCTTATCCAGAAATGGGAACT 3324
Db 243 GATCTCTATACCAAGTAATCTCTTTTGTGGAATCTTATCCAGAACTTGAGATT 302
QY 3325 CTTGGGCTAAGATATTTCTTCCTATCCGCTGATGTTGAGCTTGGAGCTGAATCCACT 3384
Db 303 CCCTGGGCAAGGTTTCTCTATCGCTTATGTTGAGATGGGTGAGAAATATAAGCA 362
QY 3385 TATCCATGCCCACTTATTCAGTGTGAGATTTCCGGAAGCCATTTGTTGAGAGACGGGGCAT 3444
Db 363 TATCCTGCTCTCTAACAGCATCAGAGGCGGAAACCTCTTTTGGAGAAATAGGACAC 422
QY 3445 ACCATCATGATCTCTTCTTGAGACTTCAGAAATACCAGTATACCTTGC 3493
Db 423 ACTATTAGCACTTACTTGTGTGACCTTCGAAATTACCAGTATACCTTGC 471

RESULT 14

US-09-983-965-4229/c

; Sequence 4229, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Rao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4229
; LENGTH: 437
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (335)
; OTHER INFORMATION:
; OTHER INFORMATION: Clone ID: 61-LIB3058-003-Q1-K1-H2
US-09-983-965-4229

Query Match 3.2%; Score 154; DB 9; Length 437;
Best Local Similarity 88.8%; Pred. No. 3e-31;
Matches 166; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 4064 GCACAGAGTGTCTTCTAGAAAACGATGACCAACTGCTAGTCTGATCTGCAG 4123
Db 379 GCACAACAGTGTCTTCTAGAGATGATGACCAACTGCTAGTCTGATCTGCAG 320
QY 4124 ATCAGAGTAGATTGACTGAGCATGTGCCAAGCTTTTGGCTCTCTGCTCTCACC 4183
Db 319 ATCAGAGCGACTAACGGAGCATGTGCCAAGCTTTTGGCTCTCTGCTCTCACC 260
QY 4184 TGAAACTTCTGAAGGAAGATGAAATGACCAACTGGGACTACGTGTGACACTGAC 4243
Db 259 TGAAGCTTCTGAAGGAAGATGAAATGACCAACTGGGACTTCGTGTGACACTGAC 200
QY 4244 ATCAGGT 4250
Db 199 ATCAGGT 193

RESULT 15

US-09-764-878-357
; Sequence 357, Application US/09764878
; Patent No. US2002009615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
; LENGTH: 118
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-357

Query Match 2.4%; Score 118; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 9.7e-22;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2058 CGATCAATGGGAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCT 2117

Db		1	CGATACAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCCT	60
Qy		2118	TGCATTAGCTCCAGATAGCCAGATAATGATCTCAGAGCTGGTCACTTTGGAATTTCT	2175
Db		61	TGCATTAGCTCCAGATAGCCAGATAATGATCTCAGAGCTGGTCACTTTGGAATTTCT	118

Search completed: May 18, 2004, 14:34:34
Job time : 1253.4 secs

this Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 18:15:22 ; Search time 399.489 Seconds
(without alignments)

15102.280 Million cell updates/sec

Title: US-09-744-167-1

Perfect score: 4839

Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1745496 seqs, 623391994 residues

Total number of hits satisfying chosen parameters: 3490992

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgm2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgm2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgm2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgm2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgm2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgm2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgm2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	318	6.6	332	6	US-10-793-479-34465
2	160	3.3	164	6	US-10-793-479-11208
3	112.8	2.3	1888	1	PCT-US03-41761-41044
4	112.8	2.3	1888	1	PCT-US03-41761-41044
5	63.8	1.3	701	5	US-09-969-034-1233
6	53	1.1	396	6	US-10-779-543-9795
7	50	1.0	2138	6	US-10-100-683-2353
8	50	1.0	2309	6	US-10-100-683-2353
9	42.6	0.9	30612	6	US-10-724-972A-326
10	42.4	0.9	3000	1	PCT-US04-07412-382
11	41.8	0.9	40335	7	US-60-568-219-6241
12	41	0.8	1286	1	PCT-US02-39555A-2125
13	40.8	0.8	1400	7	US-60-545-213-497
14	40.8	0.8	1400	7	US-60-545-213-4769
15	40.8	0.8	6129	6	US-10-791-219-10
16	40.8	0.8	6297	6	US-10-791-219-2
17	40.8	0.8	17930	7	US-60-568-845-2909
18	40.6	0.8	605	5	US-10-779-543-20366
19	40.4	0.8	67152	7	US-60-563-440-11937
20	40.2	0.8	1790242	6	US-10-767-471-10805
21	39.4	0.8	4269	6	US-10-100-683-12071
22	39.4	0.8	17386	6	US-10-681-199-10
23	39.4	0.8	308631	7	US-60-550-051-3004
24	38.8	0.8	308631	7	US-60-568-845-2891
25	38.8	0.8	393	6	US-10-779-543-11703
26	38.8	0.8	122496	7	US-60-563-440-11884

ALIGNMENTS

RESULT 1

US-10-793-479-34465

; Sequence 34465, Application US/10793479

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/10793,479

; CURRENT FILING DATE: 2004-03-03

; PRIOR APPLICATION NUMBER: US/09/513,999

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 34465

; LENGTH: 332

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-793-479-34465

Query Match 6.6%; Score 318; DB 6; Length 332;

Best Local Similarity 99.7%; Pred. No. 3.2e-77;

Matches 329; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	171	CRAGGTGTAGATGAATTTGACAAACGAAAGTGAACAGTTCTTCTACTTTATTGGA	230
DB	1	CAAGGTGTAGATGAATTTGACAAACGAAAGTGAACAGTTCTTCTACTTTATTGGA	60
QY	231	TACAAAGTGAATTAAGATTTAGATCCCTCTCTCAACGGCTGTCAATTAACCTACTTT	290
DB	61	TACAAAGTGAATTAAGATTTAGATCCCTCTCTCAACGGCTGTCAATTAACCTACTTT	120
QY	291	GGCC-AGTGTGAATGATCTGCAAGTTCTTAATGAGTCACCAACCACTGGAAGTCTCT	349
DB	121	GGCCAAAGTGTGAATGATCTGCAAGTTCTTAATGAGTCACCAACCACTGGAAGTCTCT	180
QY	350	CCCTGGCTCAATTCAGTCCCTCTGACACAGAGGAGGATCACTGTGTAATGGACAGG	409
DB	181	CCCTGGCTCAATTCAGTCCCTCTGACACAGAGGAGGATCACTGTGTAATGGACAGG	240
QY	410	ACTGTGAATTAATTCAGAGATTGCCAATGTGGATTGATGAAGTGTGTGGCAGAG	469
DB	241	ACTGTGAATTAATTCAGAGATTGCCAATGTGGATTGATGAAGTGTGTGGCAGAG	300
QY	470	ACCAGTGAATTAAGAGAACTATAGTTGGG	499


```

; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1233
; LENGTH: 701
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 517, 620, 628, 630, 634, 650, 658, 693
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-1233

```

	Query Match	1.3%;	Score 63.8;	DB 5;	Length 701;
	Best Local Similarity	66.2%;	Pred. No. 5.5e-07;		
	Matches 92;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;
Qy	2535	ACUTCACACGTGATGCTGCTGCTGGGAGTTTAAAGCACCCCTGGACACAGAGTGGCTCA	2594		
Dd	42	ACUTTCACACGACACTTTGGCAGTCTCAGACACTTAAACAACGAGGTGTTGAAGCACTATG	101		
Qy	2595	GCCACAGACAGACAGCGGAGTTTGTTTGCTGTGATGGGATCTTGCCCAATGGAGAGTTGC	2654		
Dd	102	TTCCAAAGACACAGAGAGAGTATGTTTGCGAGATGGTATATTGCCCAATGGTGAAGTTGC	161		
Qy	2655	TGATGACGCCAAATTAAACA	2673		
Dd	162	AGATACAAACAAATTATCA	180		

```

RESULT 6
US-10-779-543-9795
; Sequence 9795, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9795
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-9795

```

```
Query Match      1.1%; Score 53; DB 6; Length 396;
Best Local Similarity 59.7%; Pred. No. 0.00038;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
```

Qy	2191	ACACGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTCTCAGGCTCAAATTGCATG	2250
Db	80	AGCCCGTTGGCTGGAGAGCGCGAGTGGTCCCGACAAGAGGTGTCGGAGATGTATG	139
Qy	2251	AAATGTGAAGCCCAAGTTTACATTACACAAAGAGGAGCATCACTTCGAGAGCATGTGGGAAG	2310
Db	140	CAGTGTGACGCCAAGTTTGACTTCTCACCAGAAAGCACCACTGTGCGCGCTCGCGGAAG	199
Qy	2311	GTTTTCTGTGCTTCTGCTGTAGCCTGAA	2339
Db	200	TGCTTCTCGCACAGGTGCTGCAGCCAGAA	228

```

RESULT 7
US-10-100-683-4908
Sequence 4908, Application US/10100683
GENERAL INFORMATION:
APPLICANT: Rosen, et al.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS900
CURRENT APPLICATION NUMBER: US/10/100,683
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: US 60/040,162
PRIOR FILING DATE: 1997-03-07
PRIOR APPLICATION NUMBER: US 60/043,576
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,601
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,845
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,580
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,599
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,314
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,632
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,892
PRIOR FILING DATE: 1997-08-22
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 13468
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4908
LENGTH: 2138
TYPE: DNA
ORGANISM: Homo sapiens
US-10-100-683-4908

```

	Query Match	1.0%;	Score 50;	DB 6;	Length 2138;
	Best Local Similarity	63.1%;	Pred. No. 0.0066;		
	Matches	77;	Conservative	0;	Mismatches 45; Indels 0; Gaps 0;
Qy	2218	TGGGTACGGATCTCAGGCTCCAAATTGCATGAATGTGAAGCCAGGTTTACATTACC	2277		
Db	303	TGGGTTCCAGACCATATGGCATCACCTGCTATACCTGTGACTGTGAATCTGGTTGGCC	362		
Qy	2278	AAAAGGAGGCATCACTGCAGAGCATGTGGGAAGTTTTCTGTGTTCTCTGTGAGCTG	2337		
Db	363	AAACGAGACACCATGCAGAAATTTGGGAATGTATTTTGTGCTGGATGCTGCCACCTG	422		
Qy	2338	AA 2339			
Db	423	AA 424			

RESULT 8
US-10-100-683-2353
; Sequence 2353, Application US/10100683
; GENERAL INFORMATION:

```

; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2353
; LENGTH: 2309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13)..(13)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (652)..(652)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (677)..(677)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-100-683-2353

Query Match
Best Local Similarity 1.0%; Score 50; DB 6; Length 2309;
Matches 77; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2218 TGGGTACCGGATCTTCAGGCTCCAAATTCAGTAATGCAAGCCAGGTTTACATTCACC 2277
DB 471 TGGGTTCCAGACCATATGGCATCACACTGCTATAACTGTGACTGTGAATCTGTTGGCC 530
QY 2278 AAAAGGAGCATCATCTCAGAGCATGTGGGAGGTTTCTGTGCTTCTCTGTAGCCTG 2337
DB 531 AAAAGGAGACCATTCAGAGAAATGTGGGAATGATTTTGTGCTGTGATGCTGCCCTG 590
QY 2338 AA 2339
DB 591 AA 592

RESULT 9
US-10-724-972A-326
; Sequence 326, Application US/10724972A
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A

```

```

; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 326
; LENGTH: 30612
; TYPE: DNA
; ORGANISM: S epidermidis
US-10-724-972A-326

Query Match
Best Local Similarity 0.9%; Score 42.6; DB 6; Length 30612;
Matches 123; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1416 TCAGGACACTATGCTAGTGGAGGACAGTGTCTTGGATTGGCAGATGCAGGTCTAGA 1475
DB 26835 TGAATTTGAAATCTAATTTCTTTCACAAATTTGTCAGAAATCGTGAATCAAGATTTGA 26894
QY 1476 TTTAAAAGGAACTTGCAATTAGTGAAGTCAAGAAATGCTGATTTCTCCACTGTTATAGACAC 1535
DB 26895 TTTAAATGATGCAATGAAATCTTTTAAGAGATGAACCTTAATAATCAAGCTTTTCTGTCCA 26954
QY 1536 ACAGCAGCAATTTATCTATCTAATGTTTGTGATTCCTATGTAATGCAGACCCAGGTGT 1595
DB 26955 AGCAAGCTCAATTTATATAAATTCAGATCAAGATTTAAACAAACAAATTTGACCATGCTTT 27014
QY 1596 TTTCTTTTGTTCAAAGACTTTTACCCTCCAAAGAAAGTTCCTGCAAGAAATGTTAGATGAACAAAT 1655
DB 27015 AGTAAATGCTCGAAAGTTCCTGCAAGAAATGTTAGATGAACAAAT 27074
QY 1656 AGAGGAAAGCAAGTCAG 1672
DB 27075 TCAGGGACTCAACCAAG 27091

```

```

RESULT 10
PCT-US04-07412-382
; Sequence 382, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuhqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Kyle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Weng, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12

```

PRIOR APPLICATION NUMBER: US 10/296,115
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 10/275,027
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1920
SOFTWARE: pt_FL_genes Version 6.0
SEQ ID NO 382
LENGTH: 3000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (86)..(2008)
PCT-US04-07412-382

Query Match 0.9%; Score 42.4; DB 1; Length 3000;
Best Local Similarity 50.0%; Pred. No. 0.99;
Matches 106; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 2197 CTGGTGAGGTGGCTCCAGTATGGGTACCGGATTCAGGCTCCAAATTCATGAATGT 2256
DB 1313 CTGGGGAGAGGGCCCCACCCCTGGTGCCTGTACACACGTCATGATGTCATGAATGC 1372
QY 2257 GAAGCCAGTTTACATTCACCAAAAGAGGAGCATCTGCAGAGCATCTGGAGGTTTC 2316
DB 1373 GGCTGCACTTCTCCCTCCCTCCCTGCTGTCATCATCTGTCACGCTGTGGCAAGTCGTG 1432
QY 2317 TGTCCTTCCTGCTGCTAGCTGAATGTAACTGTATATACATGACAGAAAGGAGCTAGA 2376
DB 1433 TGCCGGAAGTTCCTCGGACACAGTACCCGCTGAAGTACCTGAAGGACAGATGGCAAG 1492
QY 2377 GTGTGTGTATCTGCCATTCAGTCTAATGAA 2408
DB 1493 GTCTGCGAGCGCTGCTCGGGAGGAGCTGAAGAA 1524

RESULT 11
US-60-568-219-6241/c
Sequence 6241, Application US/60568219
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: RESPONSIVENESS TO STATIN TREATMENT, METHODS OF DETECTION AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C0001516
CURRENT APPLICATION NUMBER: US/60/568,219
CURRENT FILING DATE: 2004-05-06
NUMBER OF SEQ ID NOS: 28217
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6241
LENGTH: 403035
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(403035)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-568-219-6241

Query Match 0.9%; Score 41.8; DB 7; Length 403035;
Best Local Similarity 55.0%; Pred. No. 23;
Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3450 CATGAATCTCTTGCAGACTTCAGAAATACCAGTATACCTTGCCAGTAGTTCAGGTTT 3509
DB 296925 CTGACTGCTTAATCTTACTTCTTTGACCCCACTGACCTTTGTTGTTATTAATAAA 296866
QY 3510 GGTGTTGATATGGAAGTTCGNAAACTAGCATCAAAATCCCAAGCAACAGATCAATGA 3569
DB 296865 TGTATTTGATATTATAATCATTTGCTGTAGCATCAAAATGCCAAGATCTTATCTA 296806
QY 3570 GATGATGAAGCCCATGACAAAGTCCAATG 3598
DB 296805 TGTAACTTAAGACAGAGTAAGCCAGTG 296777

RESULT 12
PCT-US02-39555A-2125/c
Sequence 2125, Application PC/TUS0239555A
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Jian-rui
APPLICANT: Wang, Zhiwei
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunging
APPLICANT: Ghosh, Malabika
APPLICANT: Xue, Aidong J.
APPLICANT: Asundi, Vinod
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Dunrui
APPLICANT: Goodrich, Ryle W.
APPLICANT: Chen, Rui-hong
APPLICANT: Wehrman, Tom
APPLICANT: Weng, Gezhi
APPLICANT: Xu, Chongjun
APPLICANT: Boyle, Bryan J.
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 820/PCT
CURRENT APPLICATION NUMBER: PCT/US02/39555A
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: US 60/339,739
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 10/128,558
PRIOR FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 60/365,384
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: US 60/365,091
PRIOR FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: US 60/372,615
PRIOR FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: US 60/376,045
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: US 60/372,381
PRIOR FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 3134
SOFTWARE: Pct_FL_genes Version 6.0
SEQ ID NO 2125
LENGTH: 1286
TYPE: DNA
ORGANISM: Homo sapiens
PCT-US02-39555A-2125

Query Match 0.8%; Score 41; DB 1; Length 1286;
Best Local Similarity 47.2%; Pred. No. 1.5;

	Matches	125;	Conservative	0;	Mismatches	140;	Indels	0;	Gaps	0;
Qy	2575	CTGTGAGCAGAGAAGTGGCTCTAGCC	CAGAGCAGAGAGCGAGT	TTGGTTT	GTCTGATGGGATC	2634				
Db	546	CTGTGACCAAGCTGTGGCTCTGTGGGGGCT	CCTGAGGGGGGCTGTGGCTCTGTGGGGGTT	487						
Qy	2635	TTGGCCAAATGAGAGAAGTTGCTGATGC	AGCCAAATTAACAATGAATGGAACTT	CCCTCTGCA	2694					
Db	486	CCAAAGGGGGGATGTGGTTCTTGTGGCTGCT	CCAGTCCAGCTGTGTGAAGCCCTGTGTGCT	427						
Qy	2695	GGAAACCTTGGCTGTGTGCACAGACCCAGT	CAAGCCAGTAACACAGTCTCTACAGCA	2754						
Db	426	GCTCTCAGGCTGTGGATCATCTGTCTCC	AGTCCAGTCTGCAAGCCCTGCTGCTGCT	367						
Qy	2755	GAGACGGATATTTGTCTATTCTCTGGGAGT	ATAACTCAGTGTGGAAGTCTGTGTGGAAGT	2814						
Db	366	CCTCAGCTGTGGGTGCATCTGTGCGAGT	CCAGTGTCTGTGAAGCCCTACTGCTGCCAGT	307						
Qy	2815	GCATGAATCTTATTCCCTGAAGATG	2839							
Db	306	CCAGCTGCTGTAAAGCCCTGTAGTGTG	282							

```

RESULT 13
US-60-545-213-497
; Sequence 497, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; TITLE OF INVENTION: Target Genes
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 497
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1395)..(1396)
; OTHER INFORMATION: n is a, c, g, or t
US-60-545-213-497

```

Query Match	0.8%;	Score 40.8;	DB 7;	Length 1400;
Best Local Similarity	44.6%;	Pred. No. 1.8;		
Matches 108;	Conservative 11;	Mismatches 123;	Indels 0;	Gaps 0;

QY	1488	TTGATTTAGTCAAAAGTGAAGAATGTCATTCTCCACTGTATAGACACACCAGCAGCAAA	1547
DB	1153	TGGTTGAGTGGCAGCAATGAAGTGCCTTAATAATTGTTTCATAGCCTACATCACCAG	1212
QY	1548	TTATCTATCTAAATGGTTTGATTCTCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTC	1607
DB	1213	GTCTCTGTGTCAAACCTGTGGCCACTCTATATGCACCTTGTTTACTCTTTATACAAATA	1272
QY	1608	AAAGACTTTACCTCCAAAGAGATTTCAGTAAACAGAGAAAGAAATAGAGGAAAGCAA	1667
DB	1273	ATATACTAAAGACTTTTAAARARARARARARARARARARARARARARARARARARAA	1332
QY	1668	GTCAGATGCTACTCAAAATATTTATGAACAGAGAGGGAATGAGGCCACAGAGAGGAGTGG	1727
DB	1333	AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	1392
QY	1728	AC 1729	
DB	1393	CC 1394	

RESULT 14
US-60-545-213-4769

```

: Sequence 4769. Application US/60545213
:
: GENERAL INFORMATION:
:
: APPLICANT: Wyeth
:
: APPLICANT: Mounts, William Martin
:
: TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
:
: TITLE OF INVENTION: Target Genes
:
: FILE REFERENCE: AM101083 (031896-042039)
:
: CURRENT APPLICATION NUMBER: US/60/545,213
:
: CURRENT FILING DATE: 2004-02-18
:
: NUMBER OF SEQ ID NOS: 303284
:
: SOFTWARE: PatentIn version 3.2
:
: SEQ ID NO 4769
:
: LENGTH: 1400
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: misc feature
:
: LOCATION: (1395)..(1396)
:
: OTHER INFORMATION: n is a, c, g, or t
:
: US-60-545-213-4769

```

Query Match	0.8%;	Score 40.8;	DB 7;	Length 1400;
Best Local Similarity	44.6%;	Pred. No. 1.8;		
Matches 108;	Conservative 11;	Mismatches 123;	Indels 0;	Gaps 0;
QY	1488	TTGCATTAGTGAAGTGAGAAATGTGATTTCTCCACTGTATTAGACACACAGCAGCAAA	1547	
Db	1153	TTCGTTTCAGTGGCAGCAATGAAGTTCGCAATTTAAATTTGTTTCATAGCTACATCACCAG	1212	
QY	1548	TTATCTATCTAATGGTTGTGATTCCTATGGAATGCAAGCCACAGGTGTTTCTTTTGTTC	1607	
Db	1213	GTCTCTGTGTCAACCTGTGGCCACTCTATATGCACTTTGTTTTCCTTTATTACAAATAA	1272	
QY	1608	AAAGACTTTTACCCTCCAAAGAAATTCAGCTAAACAGAGAAAAAGAAATACAGAGAAGCAA	1567	
Db	1273	ATATACTAAAGACTTTTAAARARARARAAAAAARAAAAAARRRRRRAAA	1332	
QY	1668	GTCAAGATGCTACTCAATATTTATGAAACAGAGAGGAATGAGCCACAGAGGGAGTGG	1727	
Db	1333	AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANGGGGGGG	1392	
QY	1728	AC 1729		
Db	1393	CC 1394		

```

RESULT 15
US-10-791-219-10
; Sequence 10, Application US/10791219
; GENERAL INFORMATION:
; APPLICANT: WEISMAN, LOIS
; TITLE OF INVENTION: INTRACELLULAR SIGNALING PATHWAYS IN DIABETIC SUBJECTS
; FILE REFERENCE: IOWA:048US
; CURRENT APPLICATION NUMBER: US/10/791,219
; CURRENT FILING DATE: 2004-03-02
; PRIOR APPLICATION NUMBER: 60/452,782
; PRIOR FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 6129
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-791-219-10

```

	Query Match	0.8%;	Score 40.8;	DB 6;	Length 6129;
	Best Local Similarity	55.7%;	Pred. No. 4.1;		
	Matches	78;	Conservative	0;	Mismatches 62; Indels 0; Gaps 0;
QY	2201	GTGAGGTGGCTCCAGTATGGGTACCGGATTC	CAGGCTCCAAATTCGATGAATGTGAAG	2260	
dh	445	GTGACCTGAAACAAATCTGGATCCGAGATGCA	CCATGTGAACAGTCTATGACTGTAGTG	505	

RESULT 14
US-60-545-213-4769

Qy	2261	CCAGGTTTACATTACACCAAAAGGAGGCATCACTGCAGAGCATGTGGGAAGGTTTCTGTG	2320
Db	506	AGAAATTACAACTTTAGCGCGAGACACCATTTGCCGACTAATGTCGGCAGATTTCTGCA	565
Qy	2321	CTTCCTGCTGTAGCCTGAAA	2340
Db	566	GTGCTGCTGTAATCAAGAA	585

Search completed: May 18, 2004, 13:43:32
Job time : 405.489 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:54:15 ; Search time 16461.8 Seconds
(without alignments)
17461.734 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 actccggccggggtagctc.....ttggaggcacattttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 1 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6627.2	99.9	6632	9	AB002303	AB002303 Human mRNA
2	6167.4	93.0	6280	9	BSM805726	BSM805726 Homo sapi
3	5579.8	84.1	5841	9	HSN804398	HSN804398 Homo sapi
4	4615.2	69.6	4620	9	AF434817	AF434817 Homo sapi
5	3643.8	54.9	4499	9	BC032227	BC032227 Homo sapi
6	3418.8	51.6	3793	9	AK127003	AK127003 Homo sapi
7	3337.4	50.3	5438	10	AK122241	AK122241 Mus muscu
8	3333	50.3	5480	10	BC042669	BC042669 Mus muscu
9	2655.2	40.0	5158	9	BSM809324	BSM809324 Homo sapi
10	2454.4	37.0	2902	9	BC030808	BC030808 Homo sapi
11	2254.4	34.0	252015	9	AC008382	AC008382 Homo sapi
12	2252.8	34.0	123169	9	AC008771	AC008771 Homo sapi
13	1910.4	28.8	126052	9	AC018764	AC018764 Homo sapi
14	1834	27.7	188017	2	AC146897	AC146897 Gorilla g
15	1668.8	25.2	2622	9	AK090896	AK090896 Homo sapi
16	1197.6	18.1	204269	2	AC129932	AC129932 Mus muscu
17	1149.4	17.3	110000	2	AC118331	Continuation (3 of
18	1149.4	17.3	297979	2	AC130639	AC130639 Rattus no
19	567.4	8.6	204269	2	AC129932	AC129932 Mus muscu
20	560.6	8.5	4839	9	AF104304	AF104304 Homo sapi
21	559	8.4	4769	9	BC032680	BC032680 Homo sapi
22	556.2	8.4	1977	6	AX809313	AX809313 Sequence
23	546.6	8.2	4565	9	AF130419	AF130419 Homo sapi
24	508.4	7.7	234114	2	AC111449	AC111449 Rattus no
25	500	7.5	4823	5	AF104305	AF104305 Xenopus l
26	495.8	7.5	89973	9	AL591398	AL591398 Human DNA
27	444	6.7	455	6	AX210516	AX210516 Sequence
28	318.4	4.8	533	11	G99658	G99658 S209P6519RG
29	269	4.1	405	11	G23648	G23648 human STS W
30	263.6	4.0	1237	10	BC046341	BC046341 Mus muscu
31	244	3.7	244	11	G24283	G24283 human STS W
32	238.6	3.6	887	10	BC058416	BC058416 Mus muscu
33	201.4	3.0	1444	10	BC038255	BC038255 Mus muscu
34	195	2.9	4598	3	AF239997	AF239997 Drosophil
35	193	2.9	4651	3	DME310804	AY051821 Drosophil
36	193.4	2.9	4660	3	AY051821	BD060611 Secrete
37	181.6	2.7	289	6	BD060611	AC004564 Drosophil
38	180.8	2.7	67833	3	AC004564	AC093307 Drosophil
39	180.8	2.7	179139	3	AC093307	AE003454 Drosophil
40	180.8	2.7	313634	3	AB003454	AC019884 Drosophil
41	180.6	2.7	54186	2	AC019884	AC024930 Homo sapi
42	148.4	2.2	178257	2	AC024930	AC091565 Homo sapi
43	148.4	2.2	180741	9	AC091565	AC138649 Homo sapi
44	146.8	2.2	151900	9	AC138649	AC136776 Homo sapi
45	146.2	2.2	228470	2	AC136776	

ALIGNMENTS

RESULT 1
AB002303
LOCUS Human mRNA for KIAA0305 gene, complete cds.
DEFINITION
ACCESSION AB002303
VERSION AB002303.1 GI:2224550
KEYWORDS KIAA0305.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Nagase, T., Ishikawa, K., Nakajima, D., Chira, M., Seki, N.,
Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
TITLE Prediction of the coding sequences of unidentified human genes.

Pred. No. is the number of results predicted by chance to have a

VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

JOURNAL
MEDLINE
PUBMED
REFERENCE

9205841
AUTHORS
TITLE
JOURNAL

Osamu Ohara, Kazusa DNA Research Institute;
Submitted (28-MAR-1997) Chiba 292-0812, Japan
1532-3, Yana, Kibarazu, Japan
(E-mail: cdna@kazusa.or.jp, Tel: +81-438-52-3913)

FEATURES
source

Location/Qualifiers
1. .6632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG0042"
/sex="male"
/tissue type="brain"
/clone lib="pBluescriptII SK plus"

gene

1. .6632
/gene="KIAA0305"

CDS

249. .4868
/gene="KIAA0305"
/codon_start=1
/protein_id="BAA20764.1"
/db_xref="GI:2224551"

/translation="MDSYFKAASVDLKLDDPEONDEODYLODVONAYDSNHCSVS
SELASSQTSLLPDKQCVNSCASSETSYGNTSSINEKTLKGLTSIQNEKNVIGLIDL
LSSVDGTSDEIQPLMGRCKPCLDLSMGNLVHATNEEDIKLLPDPKSNADL
LIGLDSVSTPCFVSDHSDTVREQNDTSSELQNRBIGGKELGIVDTLLSDS
YNYSGTENLKKDKIENLESIVDFNMSALTRQSKFMHAKDLQHNKSPQGLLKQV
LVKEVDVAIVTAAELCKEKGTSALCSLPKNEIDLCLNDSNRDENKLPDPSPFOED
KTVILKQSAOEDSKLNDVIOISSALHVSVDPSLCLPASGSMCGSLISBK
ARGPLPOHEKDNIDQVTHESIQNSVULGEPFKENOLLKQEKCKSLLOSLEG
MEDRKIDPDQVIAEISLDGSDTSVVEQSGELSGTHVPESDCCGPIPTSSNDM
DQGLDYFNIDEGAKSGLPDAELDAFLTEQYLQNTNKSFEENVNDSKSNQNDM
KGLDGNINNTIYFAEAGAGEGSHINIICETVDKQNTIENGSLGSKSTIPVQOGLP
TSKBITNQLSVDSIQSGVGARPKQLFSLPSTRSSKDLNKPDPDTIESEFSLAD
TWPTICADISTADPOFNSNYIDIESNSEGSSFTANEDSVKENTCKEGLVLQK
QPTWVPSSEAENCMQVFTFKRHHCRACGKVCVGCNCRKCKLOYLEKEARVCV
VCYETLSKAQAFERMPSTNSLKNHSDCTTVPQENQTSIIPSPATLPYSALKV
PVGSLGSKSKRQVAFADGILPNGEVADTKLSSGSKRCSDESPSPDPMTVNTVD
HSHSTPKNETGIDTRNEIIQSPISQVPSVEKSMNTNGELPSTGSGFTLDDVDF
AETPEPSPTGVVNSNLPSTASIDYRLDCLDKYCNKLSLLFNDEDSLPPLIVASG
EKGSPVVEHPSEHQIILLEGGFHPTVFLNANLLVNVKIFYSDDKYWFSTNG
LHGLQABIIILLCLPNEPTIPKIDIFRLFTIVKDLKGVYIENLDNIPTFESFLSS
KDHGGLFITPTFOLKDDLSLSPNPFILGILIOKLEIPWAKVPMRLMLRLGAEYKAY
PAPLTSIRGRPLRGEGHTIMLLVDLRVQVTLNIDOLLHMEGKSCIKIPRK
YSDVKNLNSNEHVISIGASPTSEADSHLVIONDGIYEQANSATGHRKVTGASF
VFNGLKTSGGFLAKSISVEDGLMVQITPETMGLRLAUREQDKFITCGKVDVLDL
KEYVDICWDAEKGNGSVISSVDGISLQGFPSKIKLEADFDTEKIVKCTEVFVL
KDQLSILSTQYQAKIEMACSAALCPHLKLSNGMKNIGLRSVSDTDMVDFQASG
BQGLLPQHYLNDLSALIPVTHGTSNSSLPLELELVFFIIEHLF"

ORIGIN

Query Match 99.9%; Score 6627.2; DB 9; Length 6632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6629; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ACTCCGCGCGGGGTAGCTCTTCACTCCTCAGCGCGAGCTGCTGTCAGATTCCCAAAAG 60
Db 1 ACTCCGCGCGGGGTAGCTCTTCACTCCTCAGCGCGAGCTGCTGTCAGATTCCCAAAAG 60
QY 61 CTCGCGAGGGCTGTAGGAGAGTGATCTCATTCATTAAACAGCTGTGTGTCGAGTTCCC 120
Db 61 CTCGCGAGGGCTGTAGGAGAGTGATCTCATTCATTAAACAGCTGTGTGTCGAGTTCCC 120
QY 121 AAATCTTTATCTATCTCAGACTTCTCTCCCTGATTCAGATTCCTATTATTCAGCTGCC 180
Db 121 AAATCTTTATCTATCTCAGACTTCTCTCCCTGATTCAGATTCCTATTATTCAGCTGCC 180
QY 181 TTGGATATCTCTCCAGAGATGTTCTCAAGGCATACAGAAATTAATTTCTGAATAGTCTG 240

181 TTGGATATCTCTCCAGAGATGTTCTCAAGGCATACAGAAATTAATTTCTGAATAGTCTG 240
QY 241 CAGGTAGGATGACAGTATTATTTAAAGCAGCTGTGTCAGTACTTTGACAAAACCTCCTTGATG 300
Db 241 CAGGTAGGATGACAGTATTATTTAAAGCAGCTGTGTCAGTACTTTGACAAAACCTCCTTGATG 300
QY 301 ATTTTGAACAGAACACAGATGAACAGATTATCTCGCAGATGTACAAATGATATGATT 360
Db 301 ATTTTGAACAGAACACAGATGAACAGATTATCTCGCAGATGTACAAATGATATGATT 360
QY 361 CTAAACCACTGCTCAGTCTTCTTCAGAGTTGGCTTCTCAGCGAACTTCTCATTTGCTCCAA 420
Db 361 CTAAACCACTGCTCAGTCTTCTTCAGAGTTGGCTTCTCAGCGAACTTCTCATTTGCTCCAA 420
QY 421 AGACCAAGAGTGGCTTAAATAGTTGTGCTCATCAGAAACAGCTATGGAACAAATGAGA 480
Db 421 AGACCAAGAGTGGCTTAAATAGTTGTGCTCATCAGAAACAGCTATGGAACAAATGAGA 480
QY 481 GTTCCCTGAATCAAAAAACACCTCAAGGGACTTACTTCTATACAAATGAAAAAATGTAA 540
Db 481 GTTCCCTGAATCAAAAAACACCTCAAGGGACTTACTTCTATACAAATGAAAAAATGTAA 540
QY 541 CAGGACTTGATCTTCTTCTTCTGATGGTGGTACTTTCAGATGAAATCCAGCGCTTAT 600
Db 541 CAGGACTTGATCTTCTTCTTCTGATGGTGGTACTTTCAGATGAAATCCAGCGCTTAT 600
QY 601 ATATGGGACGATGTAGTAACCTATCTGTGATCTCATAGTGCATGACATGGTAACTTAGTTC 660
Db 601 ATATGGGACGATGTAGTAACCTATCTGTGATCTCATAGTGCATGACATGGTAACTTAGTTC 660
QY 661 ATGCAACCAATAGTGAAGAAGATATTAATAAATAATTTATCCAGATGATTTTAAGTCTAATG 720
Db 661 ATGCAACCAATAGTGAAGAAGATATTAATAAATAATTTATCCAGATGATTTTAAGTCTAATG 720
QY 721 CAGATTCTTGTGATGGATGGATTTATCTTCAGTCTGACATGACTCCCTGTGTTCTTCAA 780
Db 721 CAGATTCTTGTGATGGATGGATTTATCTTCAGTCTGACATGACTCCCTGTGTTCTTCAA 780
QY 781 CAGACCATGATAGTATGATCTGTGACAGAAACACAGAAATGATATCAGTTCTGAAATACAAA 840
Db 781 CAGACCATGATAGTATGATCTGTGACAGAAACACAGAAATGATATCAGTTCTGAAATACAAA 840
QY 841 ATAGAGAAATCCGAGGAATCAAGAAATTTGGGTATTAAGATAGATACACACTTTTCAAGTT 900
Db 841 ATAGAGAAATCCGAGGAATCAAGAAATTTGGGTATTAAGATAGATACACACTTTTCAAGTT 900
QY 901 CCTATAATTTACAGTGGAAACAGAAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960
Db 901 CCTATAATTTACAGTGGAAACAGAAAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960
QY 961 CAATTGTTGATTTTAAACATGTCATCTGTTTGAATCTGCAAGAAATTTCCAAAATGTTTCA 1020
Db 961 CAATTGTTGATTTTAAACATGTCATCTGTTTGAATCTGCAAGAAATTTCCAAAATGTTTCA 1020
QY 1021 CCAAGAGAGCTTACACACAGAGCCAGCCATGTTGATTTAAGATGTTTGGCTTAG 1080
Db 1021 CCAAGAGAGCTTACACACAGAGCCAGCCATGTTGATTTAAGATGTTTGGCTTAG 1080
QY 1081 TAAAGAGAGAGTAGATGTGGCAGTATGAGCTTCCGCGAGAAATGTTTAAAGAGAGAGGCA 1140
Db 1081 TAAAGAGAGAGTAGATGTGGCAGTATGAGCTTCCGCGAGAAATGTTTAAAGAGAGAGGCA 1140
QY 1141 AGACAGAGTCTTTGACCTGCGCTTCCGAGAAATGAGATTTATGCTTTAATGATTTCAA 1200
Db 1141 AGACAGAGTCTTTGACCTGCGCTTCCGAGAAATGAGATTTATGCTTTAATGATTTCAA 1200
QY 1201 ATTCAAGAGATGAAATTTCAAATTTACCTGACTTTTCTCTTTTCCAGGAAGTAAAGCTGTTA 1260
Db 1201 ATTCAAGAGATGAAATTTCAAATTTACCTGACTTTTCTCTTTTCCAGGAAGTAAAGCTGTTA 1260
QY 1261 TAAACCAATCTGCAAGAGACTCAAAAAGTTTAGACTTTAAGATTAATGATTAATCC 1320

Db	1261	TAAACAATCTGCACAGAGAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCC	1320
Qy	1321	AGATTCTCTTCCAGCTTTACATGTTTCCAGTAAGAGATGCGCTCCTCATTTGTCCTGTC	1380
Db	1321	AGATTCTCTTCCAGCTTTACATGTTTCCAGTAAGAGATGCGCTCCTCATTTGTCCTGTC	1380
Qy	1381	TTCTGCTGCTGGGTCTATGTTGGATGATCAATTAATGAAAGTAAGACAGCGGGTGAATTTT	1440
Db	1381	TTCTGCTGCTGGGTCTATGTTGGATGATCAATTAATGAAAGTAAGACAGCGGGTGAATTTT	1440
Qy	1441	TACCTCAGCATGAACATAAAGATAATATACAAGATGCAGTACATATACATGAAGAAATAC	1500
Db	1441	TACCTCAGCATGAACATAAAGATAATATACAAGATGCAGTACATATACATGAAGAAATAC	1500
Qy	1501	AGAACAGTGTGTTCTAGTGGGGAAACCAFTTCAAGAGAAATGATCTTTTGAACAGGAAA	1560
Db	1501	AGAACAGTGTGTTCTAGTGGGGAAACCAFTTCAAGAGAAATGATCTTTTGAACAGGAAA	1560
Qy	1561	AATGTAAGCATACTCTTCCAGTCAATTAATGAAAGGATGGAAGACAGAAATAGATC	1620
Db	1561	AATGTAAGCATACTCTTCCAGTCAATTAATGAAAGGATGGAAGACAGAAATAGATC	1620
Qy	1621	CTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGGTGCACACAGTCTTACAGTTG	1680
Db	1621	CTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGGTGCACACAGTCTTACAGTTG	1680
Qy	1681	TAGATCTCAAGAGGGCTTTTGGCACTCATGTCCAGAGTCTTCTGATTTGTTGTAAG	1740
Db	1681	TAGATCTCAAGAGGGCTTTTGGCACTCATGTCCAGAGTCTTCTGATTTGTTGTAAG	1740
Qy	1741	GTTTTATTAATCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTACTTTAATA	1800
Db	1741	GTTTTATTAATCTTTTCAAGCAATGATATGATGGGCAAGACTTAGATTACTTTAATA	1800
Qy	1801	TTGATGAAGCGCAAAAGTGGCCCACTAATAGTGAATGCTGAACCTGATGCTCTGA	1860
Db	1801	TTGATGAAGCGCAAAAGTGGCCCACTAATAGTGAATGCTGAACCTGATGCTCTGA	1860
Qy	1861	CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAATGTAATGACTTA	1920
Db	1861	CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAATGTAATGACTTA	1920
Qy	1921	AATCGCAAAATCAATCAGATAGATGAAGGCTTAGATGATGGAACATCAATPAATATAT	1980
Db	1921	AATCGCAAAATCAATCAGATAGATGAAGGCTTAGATGATGGAACATCAATPAATATAT	1980
Qy	1981	ATTTCAATGACAGAGCAGGAGCTATTGGGAAAGTCTATGTTATTAATTAATTTGTAAG	2040
Db	1981	ATTTCAATGACAGAGCAGGAGCTATTGGGAAAGTCTATGTTATTAATTAATTTGTAAG	2040
Qy	2041	CAGTTGATAACAAAATACAAATAGAAATGGCTTTCTTTAGGAGAAAAGCACTATTTC	2100
Db	2041	CAGTTGATAACAAAATACAAATAGAAATGGCTTTCTTTAGGAGAAAAGCACTATTTC	2100
Qy	2101	CAGTTCAACAGGGTTACCTACCAAGTCTGAGATTACAAATCAATATCAGTCTCTG	2160
Db	2101	CAGTTCAACAGGGTTACCTACCAAGTCTGAGATTACAAATCAATATCAGTCTCTG	2160
Qy	2161	ATATTACAGTCAATCTGTTGGGGCCAGACCTTAAGCAATGTTTAGCCTTCCATCAA	2220
Db	2161	ATATTACAGTCAATCTGTTGGGGCCAGACCTTAAGCAATGTTTAGCCTTCCATCAA	2220
Qy	2221	GAACAAGGAGTTCAAAAGGACCTGAATTAAGCAGATGTTCCAGATACAAATAGAAAGTGAAC	2280
Db	2221	GAACAAGGAGTTCAAAAGGACCTGAATTAAGCAGATGTTCCAGATACAAATAGAAAGTGAAC	2280
Qy	2281	CCAGCACAGACAGATACCGTTGTTCCAAATCACTTGTGCTATAGATTCTACAGCTGATCCAC	2340
Db	2281	CCAGCACAGACAGATACCGTTGTTCCAAATCACTTGTGCTATAGATTCTACAGCTGATCCAC	2340
Qy	2341	AGGTTAGCTTCAACTCTAAATTAATGATAGAAAGTAAATCTGAAGGTGGAATCTAGTT	2400
Db	2341	AGGTTAGCTTCAACTCTAAATTAATGATAGAAAGTAAATCTGAAGGTGGAATCTAGTT	2400
Qy	2401	TCGTAACGTGAAATGAAGATTTCTGTACCTGAAAAACATTTGCAAAAGAGGCTTGGTTTTGG	2460
Db	2401	TCGTAACGTGAAATGAAGATTTCTGTACCTGAAAAACATTTGCAAAAGAGGCTTGGTTTTGG	2460
Qy	2461	GCCAGAAACAGCCTTACTTTGGGTTCTGATTCAGAGCTCCAACTGTATGAATGAGTCCCAAG	2520
Db	2461	GCCAGAAACAGCCTTACTTTGGGTTCTGATTCAGAGCTCCAACTGTATGAATGAGTCCCAAG	2520
Qy	2521	TCAAAATTTACTTTTACCAAAAGGGGACACCAATTCGCCAGCATGTGGGAAAGATTTTGTG	2580
Db	2521	TCAAAATTTACTTTTACCAAAAGGGGACACCAATTCGCCAGCATGTGGGAAAGATTTTGTG	2580
Qy	2581	GTGTCCTGTTGTAATAGGAAGTGTAACTGCAATATCTAGAAAAAGAACAGATATGTC	2640
Db	2581	GTGTCCTGTTGTAATAGGAAGTGTAACTGCAATATCTAGAAAAAGAACAGATATGTC	2640
Qy	2641	TAGTCTGCTATGAAGTCTTAAAGTCAAGGATTTGAAAGGATGATGATGATGATGATG	2700
Db	2641	TAGTCTGCTATGAAGTCTTAAAGTCAAGGATTTGAAAGGATGATGATGATGATGATG	2700
Qy	2701	GTCTTAATCTTAAGTCTTAATCATTTCTGATGAATGATGATGATGATGATGATGATG	2760
Db	2701	GTCTTAATCTTAAGTCTTAATCATTTCTGATGAATGATGATGATGATGATGATGATG	2760
Qy	2761	ACCAACATCCAGTATACCTTCAACAGCAACTTTGCCAGTCTCAGCATTTAAACAACAG	2820
Db	2761	ACCAACATCCAGTATACCTTCAACAGCAACTTTGCCAGTCTCAGCATTTAAACAACAG	2820
Qy	2821	GTGTTGAAGGACTATGTTTCCAAAGAACAGAGAGATGATGTTTGCAGATGATATTC	2880
Db	2821	GTGTTGAAGGACTATGTTTCCAAAGAACAGAGAGATGATGTTTGCAGATGATATTC	2880
Qy	2881	CCAATGGTGAAGTTCAGATACACAAATTTATCATCTGGAAGTAAAGATGTTCTGAAG	2940
Db	2881	CCAATGGTGAAGTTCAGATACACAAATTTATCATCTGGAAGTAAAGATGTTCTGAAG	2940
Qy	2941	ACTTTAGTCTCTCTCAGTCTGATGCTGATGACAGTAAACACAGTGGATCATTTCCCAT	3000
Db	2941	ACTTTAGTCTCTCTCAGTCTGATGCTGATGACAGTAAACACAGTGGATCATTTCCCAT	3000
Qy	3001	CTACTACAGTGAAGAACCAACAAATGACAGAGATGATTAACAGAAATGAGATAATTC	3060
Db	3001	CTACTACAGTGAAGAACCAACAAATGACAGAGATGATTAACAGAAATGAGATAATTC	3060
Qy	3061	AGAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTCTATGAAACACAGGAAATGAGG	3120
Db	3061	AGAGTCTTATTTCTCAGGTTCCATCAGTGGAAATTTCTATGAAACACAGGAAATGAGG	3120
Qy	3121	GGTTACCTACTTCTGGTTCAATTTACATGATGATGTTTTCAGAAAATGAGGAAAC	3180
Db	3121	GGTTACCTACTTCTGGTTCAATTTACATGATGATGTTTTCAGAAAATGAGGAAAC	3180
Qy	3181	CATCTAGTCTACTGCTGCTTAGTTAACAGCAATTTACCTATTGCTAGTATTTTCAGATT	3240
Db	3181	CATCTAGTCTACTGCTGCTTAGTTAACAGCAATTTACCTATTGCTAGTATTTTCAGATT	3240
Qy	3241	ATAGGTTACTGCTGATATTAACAAAGTATGCTGCAATAAGATTTAGTCTTTTACCTAATG	3300
Db	3241	ATAGGTTACTGCTGATATTAACAAAGTATGCTGCAATAAGATTTAGTCTTTTACCTAATG	3300
Qy	3301	ATGAGGACAGTTTGGCCCACTCTGTTGTCATCTGGAGAAAAGGATCAGTGCCTGTAG	3360
Db	3301	ATGAGGACAGTTTGGCCCACTCTGTTGTCATCTGGAGAAAAGGATCAGTGCCTGTAG	3360
Qy	3361	TAGAGAGACATCCATCTCATGAGCAGATCAATTTTCTTTTGAAGGTGAAGGCTTTCATC	3420
Db	3361	TAGAGAGACATCCATCTCATGAGCAGATCAATTTTCTTTTGAAGGTGAAGGCTTTCATC	3420
Qy	3421	CTGTTTACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATCCT	3480
Db	3421	CTGTTTACATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATCCT	3480

Db	5641	GTATCAATGCAATATTTTTCCTCCTTCTTCCATCAATATACCACCTGATTTACC	5700
Qy	5701	ACTTCTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTTATGTAGCAAT	5760
Db	5701	ACTTCTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTTATGTAGCAAT	5760
Qy	5761	AAATGAAGCCTGAAACAGGTTTTTTTACTTCCACTTTAAATCCTTGAAGAAATTTCTTGGCAA	5820
Db	5761	AAATGAAGCCTGAAACAGGTTTTTTTACTTCCACTTTAAATCCTTGAAGAAATTTCTTGGCAA	5820
Qy	5821	CTTCGCATATTTTCATTCACATCGTGTATAGTATATAAATTTAAATGAACATAATCTTT	5880
Db	5821	CTTCGCATATTTTCATTCACATCGTGTATAGTATATAAATTTAAATGAACATAATCTTT	5880
Qy	5881	TGCATATTTAAATCTTTTATATAGTATATTTTATACAGGATATTAACATAAGTT	5940
Db	5881	TGCATATTTAAATCTTTTATATAGTATATTTTATACAGGATATTAACATAAGTT	5940
Qy	5941	AAATCCCTATGTTTAAATGTTTACAGAGCTTTCTCTTACTTCAACAGCAAAAAG	6000
Db	5941	AAATCCCTATGTTTAAATGTTTACAGAGCTTTCTCTTACTTCAACAGCAAAAAG	6000
Qy	6001	TGGGGGGCATATTTAGTCTCTGTCATTTAAGTATGTAAGAAATTTAATCATTATTTTGA	6060
Db	6001	TGGGGGGCATATTTAGTCTCTGTCATTTAAGTATGTAAGAAATTTAATCATTATTTTGA	6060
Qy	6061	TGCTTTAAACATCTCATGTATATATATGTTTGTATCAAAAACACTCATATTTTCA	6120
Db	6061	TGCTTTAAACATCTCATGTATATATGTTTGTATCAAAAACACTCATATTTTCA	6120
Qy	6121	AGAAAAAGAAATTAATGTTAAATAGCCCTGTTTAAAGAAATATTTATGAAGCATCTCAA	6180
Db	6121	AGAAAAAGAAATTAATGTTAAATAGCCCTGTTTAAAGAAATATTTATGAAGCATCTCAA	6180
Qy	6181	CTTGACATCAAGTCAAGTATTAATCAAGCTGAGGCTCTCAAGCTAGGAGACTGA	6240
Db	6181	CTTGACATCAAGTCAAGTATTAATCAAGCTGAGGCTCTCAAGCTAGGAGACTGA	6240
Qy	6241	GAATTTTAAATCAGTTTCGGCATATAGTTTGGCATGATCATCTGTAGTACTTAGCCAA	6300
Db	6241	GAATTTTAAATCAGTTTCGGCATATAGTTTGGCATGATCATCTGTAGTACTTAGCCAA	6300
Qy	6301	AGACAAATTCGAGAGAAATATCAGCCTTCGGAAGTAGCTCTCCTGAACATGTAAG	6360
Db	6301	AGACAAATTCGAGAGAAATATCAGCCTTCGGAAGTAGCTCTCCTGAACATGTAAG	6360
Qy	6361	TGTCGAGATATCAATAAATGCAACCTGTTTATAATTTGTGAATTTATTCGAATGTT	6420
Db	6361	TGTCGAGATATCAATAAATGCAACCTGTTTATAATTTGTGAATTTATTCGAATGTT	6420
Qy	6421	GTAAGATGAAACAAATTCATATCAAAACCAATTTATGTTTCTTAATATAGTATGTA	6480
Db	6421	GTAAGATGAAACAAATTCATATCAAAACCAATTTATGTTTCTTAATATAGTATGTA	6480
Qy	6481	TTCTGCCATGTAAGTAAATGCAAGCTCTTAAATTAACCAATGTTAGAGGCTGTTCCAT	6540
Db	6481	TTCTGCCATGTAAGTAAATGCAAGCTCTTAAATTAACCAATGTTAGAGGCTGTTCCAT	6540
Qy	6541	GATGGGAGAGCTTTGGATTTGTTTTCATAAATCTTACATTCATATAAATTTGGAATTA	6600
Db	6541	GATGGGAGAGCTTTGGATTTGTTTTCATAAATCTTACATTCATATAAATTTGGAATTA	6600
Qy	6601	TGTCCTGAAATTTGGAGGCAATTTTGAAGT	6632
Db	6601	TGTCCTGAAATTTGGAGGCAATTTTGAAGT	6632

RESULT 2

HSM805726 HSM805726 6280 bp mRNA linear PRI 16-JUN-2003
 LOCUS Homo sapiens mRNA; cDNA DKFZp686K236 (from clone DKFZp686K236);
 DEFINITION complete cds.
 ACCESSION BX537424

BX537424.1 GI:31873349

VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 6280)

Ansong, H. W., Krieger, S., Regiert, T., Rittmüller, C., Schwager, B.,
 Newes, H. W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
 Wiemann, S.

TITLE
JOURNAL

Direct Submission

Submitted (16-JUN-2003) M2FS, Ingolstaedter Landstr.1, D-85764

COMMENT

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp686K236) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at http://mips.gsf.de/proj/cDNA/.

FEATURES
source

1..6280
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /map="5q14.1"
 /clone="DKFZp686K236"
 /tissue_type="human cervix"
 /clone_lib="686 (synonym: hicc3). Vector pSport1_Sfi; host
 DH10B; sites SfiIA + SfiIB"
 /dev_stage="adult"

gene

1..6280

/gene="DKFZp686K236"

CDS

133..4752

/gene="DKFZp686K236"

/note="endofin"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAD97666.1"

/db_xref="GI:31873350"

/translation="MDSYFKAASDLKLLDDFEONPEQDYLDQVQAYDSNHCVS

SELASQRTSLPDKQECVNCASSETSYGNESLNKTKLGLTISQNKXVTLGLDL

LSSVDGYSDEIPLYMRCKSPICDLIDMGNLVHATNSEEIDKLLPDDPKSNADS

LIGLSSVSTPCVSTSDHSDTVREQNDTSSELRNREIGIKELGKVDITLSDS

VYSGTENLKKIKFQLESIVDFNMSSALTQSSKMFHAKLKHQSPQCLLKDVG

LVKEVDVAVITAAELKEEGTSALTCTLPKNECLDCLNSRDENFKLPDFSFQED

KTVIKSAQEDSKSLDKNDVIODSSALHVSCKDVPSSLCLPAGSGMSGLTESK

AKGDLPOHEKNDIQDVTIHEEQNSVVLGGEPFENDLLKQCKSKILLQSLIEG

MEDRKIDPQTVIRAESLDGDTSTVVEQGLSGTHVPSSDCCGFINTPSSNDM

DQDDLYFNIDEGAKSGPLIADAEADFLTYQLQNTIKGFENNDSKQMNQIDM

KGLDDGNINITYFNAEAGIAGESHGINIICETVDKONTIENGSLGKSTIPVOOGLP

TSKSRITNOLSDINSQSGARQKLPSPRTRSSKDLNKPDPVPTIIESEPSTAD

QTVTPYDSEAPNCMMCOVKFTFKERHCRAGCKVFCVCCNCKCKLOYLEKARVCV

VCYETLSKRAFERMMSGTGSLKSNDECTTVQPPENQTSSTPSPATLPSALKQ

POVEGCKEQRKRVMPADGILPNGEADTTLKSSGKRCSEDPSLSPDVTNIVD

HSHTVTKPNNETGDIITRNEIISQIPVSPVEKLSMKNTEGNETSGSTLDDDFV

ABTEPSSPTGLVNSNPIIASISQIRLLCINKYCNKISLLPNDESLPELLVASG

BKSGVPVBEHPHQIILLLEGEPHPTVFLNANLVNKKVFIYSSDKWYFTSTNG

LHGLGQAEIILLCLPNEDTIPKDIIFRITFYKALKGKYIENLDNITPSTFLSS

KDHGSLFTPTFKLDDLSLPNPLGILIQLEIIPKVPNRLMLRGAAYKAY

PAPLTSIRGRKFLPGEIHTINLLVLDNRYQVTLHNIDQLLHNMKSKCIKPRKK

YSDVMKLVNSNEHVISIGASFTEADSHLVCIQNDGIYEQTQANSTIGPRKVTGAS

VYFGALTKTSGFLAKNSIVEDGLMVQITPTMNGRLRLALEQDFKTCGKVTADVL

REYVDICWDAEENKNGVISVDGILSQGFPSEKILEADFETDEKIVCTEVPYFL

KQDQLSILTSYQFAKEIAMACSAALCPHLTKSGNMKIGLRVSIIDTMVFEQAGS

EQQLLPQHYLNDLSALLPVIHVGTSNSSLPLEIELVFFIEHLF"

Query Match

93.0%; Score 6167.4; DB 9; Length 6280;

QY 2345 TAGCTTCAACTCTAAATACATTTGATATAGAAAGTAATTCCTGAAGTGGATCTAGTTTCGT 2404
Db 2229 TAGCTTCACTCTAAATACATTTGATATAGAAAGTAATTCCTGAAGTGGATCTAGTTTCGT 2288
QY 2405 AACTGCAAAATGAAGATTCCTGATCTGAAACACATTCGAAAGAGGCTTGTGTTGGCCCA 2464
Db 2289 AACTGCAAAATGAAGATTCCTGATCTGAAACACATTCGAAAGAGGCTTGTGTTGGCCCA 2348
QY 2465 GAAACAGCTACTTGGGTTCCTGATTCAGAAAGTCCAAATCTGATGAACCTGCCAAGTCAA 2524
Db 2349 GAAACAGCTACTTGGGTTCCTGATTCAGAAAGTCCAAATCTGATGAACCTGCCAAGTCAA 2408
QY 2525 ATTTACTTTTACCAAAACGGGACACCAATTCGAGAGCATGTGGGAAAGTATTTTGTGGTGT 2584
Db 2409 ATTTACTTTTACCAAAACGGGACACCAATTCGAGAGCATGTGGGAAAGTATTTTGTGGTGT 2468
QY 2585 CTGTTGTAATAGGAAGTGAACCTGCAATATCTAGAAAGAGGACAGAGATGATGTGATGT 2644
Db 2469 CTGTTGTAATAGGAAGTGAACCTGCAATATCTAGAAAGAGGACAGAGATGATGTGATGT 2528
QY 2645 CTGCTATGAACCTATTAGTAAAGCTCAGGCAATTTGAAAGGATGATGATGCCAAGTCTC 2704
Db 2529 CTGCTATGAACCTATTAGTAAAGCTCAGGCAATTTGAAAGGATGATGATGCCAAGTCTC 2588
QY 2705 TAATCTTAAGTCTTAATCAATTCGTATGATGATGATGATGATGATGATGATGATGATGATG 2764
Db 2589 TAATCTTAAGTCTTAATCAATTCGTATGATGATGATGATGATGATGATGATGATGATGATG 2648
QY 2765 AACATCAGATATACCTTCACAGCAACTTTGCCAGTCTCAGCACTTAAACAAACAGAGTGT 2824
Db 2649 AACATCAGATATACCTTCACAGCAACTTTGCCAGTCTCAGCACTTAAACAAACAGAGTGT 2708
QY 2825 TGAAGGACTATGTTTCCAAAGAACAGAGAGATGATGATGATGATGATGATGATGATGATGATG 2884
Db 2709 TGAAGGACTATGTTTCCAAAGAACAGAGAGATGATGATGATGATGATGATGATGATGATGATG 2768
QY 2885 TGGTGAAGTTCAGATACAAACAAATTCATCTCGGAAGTAAAGATGATGATGATGATGATGATG 2944
Db 2769 TGGTGAAGTTCAGATACAAACAAATTCATCTCGGAAGTAAAGATGATGATGATGATGATGATG 2828
QY 2945 TAGTCTCTCTCAGCTGATGTCCTATGACAGTAAACACAGTGGATCATTCCTCCATCTAC 3004
Db 2829 TAGTCTCTCTCAGCTGATGTCCTATGACAGTAAACACAGTGGATCATTCCTCCATCTAC 2888
QY 3005 TACAGTGAAGAACCAATGACAGAGAGATATACAGAAATGAGATGATGATGATGATGATGATG 3064
Db 2889 TACAGTGAAGAACCAATGACAGAGAGATATACAGAAATGAGATGATGATGATGATGATGATG 2948
QY 3065 TCCTATTTCTCAGGTTCCATCAGTGGAAATTTGCTATGAACACAGGAAATGAGGGGTT 3124
Db 2949 TCCTATTTCTCAGGTTCCATCAGTGGAAATTTGCTATGAACACAGGAAATGAGGGGTT 3008
QY 3125 ACCTACTTCTGGTTCATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3184
Db 3009 ACCTACTTCTGGTTCATTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 3068
QY 3185 TAGTCTCTCTGCTGCTTGTAGTTTAAACAGCAATTTACCTATGCTAGTATTTTACAGATTTAG 3244
Db 3069 TAGTCTCTCTGCTGCTTGTAGTTTAAACAGCAATTTACCTATGCTAGTATTTTACAGATTTAG 3128
QY 3245 GTTACTGTGTGATATTAACAAGTATGCTGCAATAAGATGATGCTTCTACCTAATGATGA 3304
Db 3129 GTTACTGTGTGATATTAACAAGTATGCTGCAATAAGATGATGCTTCTACCTAATGATGA 3188
QY 3305 GGACAGTTTGGCCCACTCTGGTTGCTATCTGGAGAAAGGATCAGTGCCTGTAGTAGA 3364
Db 3189 GGACAGTTTGGCCCACTCTGGTTGCTATCTGGAGAAAGGATCAGTGCCTGTAGTAGA 3248
QY 3365 AGAACATCCATCTCATGACAGATCAATTTTGTCTTCTGAAGGTGAAGGCTTTCATCTCTGT 3424
Db 3249 AGAACATCCATCTCATGACAGATCAATTTTGTCTTCTGAAGGTGAAGGCTTTCATCTCTGT 3308

QY 3425 TACATTTGTCTAAATGCTAAATCTACTCTGTAATGTCAAAATTCATATTTTATTTCTCTCAGA 3484
Db 3309 TACATTTGTCTAAATGCTAAATCTACTCTGTAATGTCAAAATTCATATTTTATTTCTCTCAGA 3368
QY 3485 CAAATATTGGTACTTTTCAACCAATGGATTCATCGCTTGGACAGGACAGAAATATTATAT 3544
Db 3369 CAAATATTGGTACTTTTCAACCAATGGATTCATCGCTTGGACAGGACAGAAATATTATAT 3428
QY 3545 TCTATTGTATGTTTTCGCAAAATGGAAGATCTATTCCTTAAGGACATCTTCAGACTATTATAT 3604
Db 3429 TCTATTGTATGTTTTCGCAAAATGGAAGATCTATTCCTTAAGGACATCTTCAGACTATTATAT 3488
QY 3605 CACCATATATAGGATGCTCTAAAGAGGAAATACATAGAAACTTTGGACAAATATTACCTT 3664
Db 3489 CACCATATATAGGATGCTCTAAAGAGGAAATACATAGAAACTTTGGACAAATATTACCTT 3548
QY 3665 TACTGAGAGTTCCTCAGTAGCAAGATCACGGAGATTCCTGTTTATTATPACACCTACTTT 3724
Db 3549 TACTGAGAGTTCCTCAGTAGCAAGATCACGGAGATTCCTGTTTATTATPACACCTACTTT 3608
QY 3725 TCAGAACTTGTATGATCTCTCATTACCAAGTAACTCTTTCTTCTGTTGGAATTTCTTATCCA 3784
Db 3609 TCAGAACTTGTATGATCTCTCATTACCAAGTAACTCTTTCTTCTGTTGGAATTTCTTATCCA 3668
QY 3785 GAAGCTTGAGATTCCTCGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATTCGGTGC 3844
Db 3669 GAAGCTTGAGATTCCTCGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATTCGGTGC 3728
QY 3845 AGAATATAAGCATATCTGCTCCTTAAACAGCATCAGAGCCGGAACCTCTTTTGG 3904
Db 3729 AGAATATAAGCATATCTGCTCCTTAAACAGCATCAGAGCCGGAACCTCTTTTGG 3788
QY 3905 AGAATATAGACACACTATTATGAACCTTACTCTGTTGACCTTCGAAATTCACGATATACCTT 3964
Db 3789 AGAATATAGACACACTATTATGAACCTTACTCTGTTGACCTTCGAAATTCACGATATACCTT 3848
QY 3965 GCATATATAGATCAACTCTTGATTCATATGAAATGGAAAGCTGCATATAAATATCC 4024
Db 3849 GCATATATAGATCAACTCTTGATTCATATGAAATGGAAAGCTGCATATAAATATCC 3908
QY 4025 ACGGAAATAGTACAGTATGTAATGAAGTACTTAATCTTCCATGACATGCTCATTAG 4084
Db 3909 ACGGAAATAGTACAGTATGTAATGAAGTACTTAATCTTCCATGACATGCTCATTAG 3968
QY 4085 CATTGAGCAAGTTTCAGTACAGAGCAGATTCCTCATCTGCTGTATACAGAAATGATGG 4144
Db 3969 CATTGAGCAAGTTTCAGTACAGAGCAGATTCCTCATCTAGTCTGTATACAGAAATGATGG 4028
QY 4145 AATTATGAAAACAGGCCCAACAGTCCACTGGCCATCTAGAAAGTGCACAGGTCAAG 4204
Db 4029 AATTATGAAAACAGGCCCAACAGTCCACTGGCCATCTAGAAAGTGCACAGGTCAAG 4088
QY 4205 TTTTGTGGTATTCAATGGAGCTCTAAACACATCTTCAGGATTTCTTGTAAAGTCCAGCAT 4264
Db 4089 TTTTGTGGTATTCAATGGAGCTCTAAACACATCTTCAGGATTTCTTGTAAAGTCCAGCAT 4148
QY 4265 AGTTGAAGTGGCTTAAATGTTACAAATAAATCTCAGAGACCATGAATGGCTTGGCGGTAGC 4324
Db 4149 AGTTGAAGTGGCTTAAATGTTACAAATAAATCTCAGAGACCATGAATGGCTTGGCGGTAGC 4208
QY 4325 TTTTACGAGACAGAAAGACTTTTAAATTTACATGTTGGGAAAGTTGATGCTAGTACACCTGAG 4384
Db 4209 TTTTACGAGACAGAAAGACTTTTAAATTTACATGTTGGGAAAGTTGATGCTAGTACACCTGAG 4268
QY 4385 AGAATACGTTGGATATCTGCTGGGTAGATGCTGAAGAAAAAGGAAAAAAGGAGGTATTCAG 4444
Db 4269 AGAATACGTTGGATATCTGCTGGGTAGATGCTGAAGAAAAAGGAAAAAAGGAGGTATTCAG 4328
QY 4445 TTCAAGTGGATGAATATCATTTCAAGGATTTCCAGTGAAGAAATAAACTTGAAGCAGA 4504
Db 4329 TTCAAGTGGATGAATATCATTTCAAGGATTTCCAGTGAAGAAATAAACTTGAAGCAGA 4388
QY 4505 TTTTGAACCGATGAGAGATTTGTAATATGATCCGAGGTGTTCTTACTTTCTTAAAGGACCA 4564

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DFKZ) ; Email: s.wiemann@dkfz-heidelberg.de; sequenced by GSF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DFKZ451M1519) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mibs.gsf.de/proj/cDNA/>.

FEATURES

Location/Qualifiers

source

1: 5841

/organism="Homo sapiens"

```
/mol type="mrna"
```

```
/db xref="BZPD:DKFZp451M1519"
```

/db_xref="taxon:960

/map="5014.1"

```
/map= 3914.1  
/clone="PKFZn451M1519"
```

```
/clone= DNF2P43M1315
/tissue type="human spinal cord"
```

```
/clone_type=human:spinal_cord
/clone_title="457 (synonym: h1cc1) vector nsport1: host
```

```
/clone_lib="451" (synonym:
DH10B: sites NotT + SalT"
```

CHILD; sizes NOT/

```

/dev/sg1
1 5841

```

L: .5841
/gene="DNKF7D451M1519"

3847 /gene="UK"

<1. .384/
/gene="DNKEZD451M1519"

```
/gene="DKFZp451ML1519"
```

```

/note="endoin, N-terminus trunc
/codon start=2

```

```
/codon_start=2
/sequence="h:acth:acj
acth:acj"
```

```
/product="hypothetical"
```

/protein_id="CAD89968":

/db_xref="GI:302683376"
/translation="KKDLGKQPGCLLQVLGVKKEEVDAVIAAECKKEGKTSA
TCSLSPKNEIDCLNDSRDNELPKPDFQEDKTVIKQSAQESKSLDLKNDVIOOS
SSALHYSKQDPVSSLSCLPSGCMGSLIKESKARDGLPQHEKNDIKQAVITHEBIO
NSVVLGPEFKENDILKQBKSKLIQSLIKEMBRKIDPDQTVIRAESLVDGDTGST
VFSQGLSAGHVPSDSCCEGFIITFSNMDQGDIDYVINDGKSGGLPIDASDL
AFTTEYLOCTWIKSPENVDNSKMQOIMDKGLDGNINLWVFNABAGIAGESHKI
NIICEVDQNTIENLSGRKSLIPVQOGLPIKSKBITWQVSLVINDSOSVGARPK
NICFSPKRSKLDNKLDPDITISEBSETADVPVTCALDSTADGPFVSNFYIDPI
HNSRGGSGFVIANEDSPENTCKRGVLQKQGPWPDSEKAPMNCMVQKFTFKR
EHCACGKFCVGNCRKQGLQJLKEARVCWCVETYSKQAFERMPSTGSLNKL
HSDPETQPOBENOTSPISATPILPSAALQKQVGLGKSCRFKWFADGLPNGEV
ADDTLLSGSKGSEDFSPLSPODMVNTVDHSHSTVSKPKNETGIDTRBISIQS
ISQVPSVEKLSMNTGNEGLTSGSTFLDDVPASTSPSPGTVGLNSNLIPIASIDY
RLDLCKINQVCKNLSILPNDEDSLPLVASGSESPGVVEEHPGQIILLEGSGF
HPVTVFLNALKLVKYEIYSSDKYWYESTWGLHQLGCAIITLLCLPNEDIPKDI
FLGFIITKDALIKYKNIENLNIITTESFASKYKQDGLGFLFITPTFKDLDSLSPSNPF
LCLITLOKUEIPWAKVFWMLRLNLSKAPPLATISIRGRKLPCEIGHTIMNLLV
DLRNYQTLDHIDOLLHWMGSKCIKIPRKYISGVNKLNSNHEVVISIGSFTEA
DSHLVQIONDGYTOANSATGHPKRVKSGFVFNAGKLTSGFSAKSIVEDFGL
QITPTSGFKRLRLRQKQFKITCKGVDAVRLRYDITCWYDAEENKGVGISVSDGI
SLOTPFSEKILAEADPEDEBKIVTEYFVLKQDLSLSTYSQFAKELTAMACSAAL
NPKLTKLNGNMNKLISRYVISITDVMVBFQAGSEGLLIPQHYLNDLSALIPIVHGCTS
CPSLPLEIETAPILGRIYIP#

ORIGIN

Query Match 84.1%: Score 5579.8: DB 9: Length 5841:

Query Match	93.1%	Score	33738
Best Local Similarity	99.8%	Pred No.	0

Best local similarity 99.8%, freq: NO: 0,
Matches 5598: Conservative 0: Mismatches 12: Indels 1: Gaps 1:

Qy 1022 CAAGACAAGCTTACACACAAGAGCCAGCCATGTGGATTACTAAAGAATGTTGGCTTAGT 1081

Db 1 CAAGACAAGCTTACACACAAGAGCCAGCCATGTGGATTACTAAAGAATGTTGGCTTAGT 60

Qy 1082 AAAAGGGAAC TAGATGTGGCAGCTATAACTGCCGCAGAAATGTTTAAAGAGAGAGGGCAA 1141

Dβ 61 AAAAGGGAAGTAGATGTGGCAGCTATAACTGCCGCAGAAATGTTTAAAGAGAGAGGGCAA 120

QY 1142 GACAAGTCTTTGACCTGCAGCCTTCGAAAATGAAGATTATGCTTAAATGATCAAA 1201

DB 121 GACAAGTCTTTGACCTGCAGCCTTCGAAAATGAAGATTATGCTTAAATGATCAAA 180

1202 TTCAAGAGATGAAAAATTTCAAATTACCTGACTTTTCTCTTTCAGGAAGATAAGACTGTTAT 1261

Db 1261 CAGCAGCAGATACCGTGTGTTCCAACTCAGTGTGCTATAGATTTCTACAGCTGATCCACA 1320
Qy 2342 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2401
Db 1321 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1380
Qy 2402 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2461
Db 1381 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1440
Qy 2462 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2521
Db 1441 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1500
Qy 2522 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2581
Db 1501 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1560
Qy 2582 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2641
Db 1561 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1620
Qy 2642 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2701
Db 1621 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1680
Qy 2702 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2761
Db 1681 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1740
Qy 2762 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2821
Db 1741 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1800
Qy 2822 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2881
Db 1801 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1860
Qy 2882 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2941
Db 1861 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1920
Qy 2942 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3001
Db 1921 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 1980
Qy 3002 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3061
Db 1981 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2040
Qy 3062 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3121
Db 2041 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2100
Qy 3122 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3181
Db 2101 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2160
Qy 3182 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3241
Db 2161 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2220
Qy 3242 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3301
Db 2221 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2280
Qy 3302 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3361
Db 2281 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2340
Qy 3362 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 3421
Db 2341 GGTAGCTTCAACTCTAAATACATGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTT 2400

Qy 3422 TGTACATTTTGTCTTAAGTCTAATCTACTCGTGAATGTCAAAATTCATATTTTATTCCTC 3481
Db 2401 TGTACATTTTGTCTTAAGTCTAATCTACTCGTGAATGTCAAAATTCATATTTTATTCCTC 2460
Qy 3482 AGACAAATATTTGTTTCAACCAATGGAATGCTTTGGGACAGGACAGAAATTTAT 3541
Db 2461 AGACAAATATTTGTTTCAACCAATGGAATGCTTTGGGACAGGACAGAAATTTAT 2520
Qy 3542 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3601
Db 2521 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2580
Qy 3602 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3661
Db 2581 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2640
Qy 3662 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3721
Db 2641 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2700
Qy 3722 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3781
Db 2701 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2760
Qy 3782 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3841
Db 2761 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2820
Qy 3842 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3901
Db 2821 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2880
Qy 3902 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3961
Db 2881 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 2940
Qy 3962 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4021
Db 2941 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3000
Qy 4022 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4081
Db 3001 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3060
Qy 4082 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4141
Db 3061 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3120
Qy 4142 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4201
Db 3121 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3180
Qy 4202 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4261
Db 3181 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3240
Qy 4262 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4321
Db 3241 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3300
Qy 4322 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4381
Db 3301 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3360
Qy 4382 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4441
Db 3361 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3420
Qy 4442 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 4501
Db 3421 TATTCATTTGTTTATGTTTCCCAATGAGATGCTTCTTAAGGACATCTTCAGACTATT 3480

AF434817 LOCUS AF434817 4620 bp mRNA linear PRI 12-NOV-2001
 DEFINITION Homo sapiens endofin mRNA, complete cds.
 ACCESSION AF434817
 VERSION AF434817.1 GI:16904135
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 4620)
 RefSeq: L.F. and Hong, W.
 Endofin, an endosomal FYVE domain protein
 J. Biol. Chem. (2002) In press
 REFERENCE 2 (bases 1 to 4620)
 RefSeq: L.F. and Hong, W.
 Direct Submission
 TITLE Submitted (17-OCT-2001) Membrane Biology Laboratory, Institute of
 Mol & Cell Biol, 30 Medical Drive, Singapore 117609, Singapore
 FEATURES
 Location/Qualifiers
 1..4620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1..4620
 /note="endosomal FYVE domain protein"
 /codon_start=1
 /product="endofin"
 /protein_id="AAL30772.1"
 /db_xref="GI:16904135"
 /translation="MDSYFKAASVLDLKLDDPFDPEODYLDQVQNAVDSNHCVS
 SELASSORTSLPKDQSCVNSCASSETSYGNESLSNEKTLKGLTSIQNEKNVTGLDL
 LSSVDGSDIQLYWRGSKPICDLISMGNLVHATNSEEDIKLLPDPFKSNADS
 LIGLDSVSDPTSSDTHDSPTLRQQNDTSSELQNEIGGKELGILGIVTILSDS
 YNGSTENLADKIFNLEISIVDFNMSALTQSSKMFHAKDLQHSQCGLLKDVG
 LVKEEDVAVITAECLKEGKTSALTSIPKQEDLCNDSNRDENFKLPDFSFED
 KTVIKQAQOEHKSLDKNDVIOSSSALHVSFKDPSLSCLPASGMSGSLIESK
 ARGDFLPQHEHDKNOADVTIHEIIONSVLGGPPKENDLLKQEKCKSLLOSLLEG
 MEDKIPDQNVPAESLDGDSSTVSVEQGLSGTHPVESDSCCEGFINTFSNDM
 DQQLDIPNDEGAKSPLSDAELDAFLTEYQLQTNIKSPFENVYNDKSNQNDM
 KGLDDGNINNIYFAEAGAGESGINIICETVDQKNTIENGSLSEKSTIPVQQSLP
 TSSEITINQSVSDINSQSVGARPDKFLSPGTRSSDLNKPDPDTLIESPFDAD
 TVPITTCADISTADPOVSFNSYDIESNBSGSSFTANEDSPVENTCKEGLVQK
 QTVWVPDSEAPNCMCOVKFTFKRRHRCACGVFCVCCNCKLOYLEKEARVCV
 VCYETISKAQAFERMSPTGSLNKSNDCTVQPOENOTSIPSPATLPVSAKQ
 PGVGLSKQEKRYWFDAGILPGEVADTTKLSGSKRCSDFSPSPDPMVTNTVD
 HSHSTVEKPNNEGTITREIITQSPISQVPSVEKLSMNTGNEGLPTSGSFLLDDVF
 AETEEPSPTGVNLSNLPASISDYELLCIDINKYCNKISLLPNDSDSLPPLVASG
 EKGSPVVEHPSEHQIILLLEGEPFVTVFNALLVNVKFIYSSDKYWFYFSSNG
 LHGLQABIIILLCLPNEITPKDIPLRFLITKDALKGKIENLDNITFTFESFLS
 KDHGFLPTFTKQDLDLSPNPLFCGLIKLEIPWAKFPFMRMLRLGLAEYKAY
 PAPLTSIRGRPLFGEIGTITMNLVLRNVQVTLNIDOLLHMEMGSKCIKPKK
 YSDVMKVLNSNEHVIIGASFTSEADSHLVCLONGIYETOANSATGHPKVTGAS
 VVFGALKTSGLFLAKSIVVEDGLMWQITPETHNGLRLAUREQDKFITCGKVDVLD
 REYVDICWDVAEKNKSVISVDGTSIQGFPEKIKLEADFTSDKIVKCTSVFVFL
 KDQLSLTSYQPAKSIAMWASALCPHLKLSNGMKNIGLRLVSDITDWFQFQGS
 EGQLLPQHYLNDLSALIPVHGTGNSLSPLFELVFFIIEHLF"

ORIGIN

Query Match 69.6%; Score 4615.2; DB 9; Length 4620;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4617; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 249 ATGACAGTATTTAAAGCAGCTGTGACGTGACCTGGACAACTCTTGTGATGTTTGA 308
 1 ATGACAGTATTTAAAGCAGCTGTGACGTGACCTGGACAACTCTTGTGATGTTTGA 60
 309 CAGAACCCAGATGAACAGATTATCTGCAGATGTACAAATGCATGATGTTCTAACCA 368
 61 CAGAACCCAGATGAACAGATTATCTGCAGATGTACAAATGCATGATGTTCTAACCA 120
 369 TGCTCAGTTCCTTCAGAGTTGGCTTCCTCAGAGGCAACTCAITGCTCCCAAGACCAA 428

121 TGCTCAGTTCCTTCAGAGTTGGCTTCCTCAGAGGCAACTCAITGCTCCCAAGACCAA 180
 429 GAGTGGTGTAAATAGTTGTGGCTTCATCAGAAAACAAGCTATGGAACAAATGAGAGTTCCCTG 488
 181 GAGTGGTGTAAATAGTTGTGGCTTCATCAGAAAACAAGCTATGGAACAAATGAGAGTTCCCTG 240
 489 AATGAAAAAACAACACTCAAGGACTTACTTCTATACAAAATGAAAAATGTAACAGGACTT 548
 241 AATGAAAAAACAACACTCAAGGACTTACTTCTATACAAAATGAAAAATGTAACAGGACTT 300
 549 GATCTTCTTCTTCTGTGGTGGTGGTACTTCCAGATGAAATCCAGCCGCTTATATATGGA 608
 301 GATCTTCTTCTTCTGTGGTGGTGGTACTTCCAGATGAAATCCAGCCGCTTATATATGGA 360
 609 CGATGTAGTAAACCTATCTGTGATCTGATAAGTGCATCGGGTAACTTAGTTCATGCAACC 668
 361 CGATGTAGTAAACCTATCTGTGATCTGATAAGTGCATCGGGTAACTTAGTTCATGCAACC 420
 669 AATAGTGAAGAGATATTAAAAAATTATTCAGATGATTTTAACTCTAAATGCAGATTCC 728
 421 AATAGTGAAGAGATATTAAAAAATTATTCAGATGATTTTAACTCTAAATGCAGATTCC 480
 729 TTGATTGGATTGGATTATCTTTCAGTGTGCAGATCTCCCTGTGTTTCTTCAACAGACCAT 788
 481 TTGATTGGATTGGATTATCTTTCAGTGTGCAGATCTCCCTGTGTTTCTTCAACAGACCAT 540
 789 GATAGTGATCTGTGCAGAACAAAGATGATATCAGTCTCTGATTTACAAATAGAGAA 848
 541 GATAGTGATCTGTGCAGAACAAAGATGATATCAGTCTCTGATTTACAAATAGAGAA 600
 849 ATCGGAGGAATCAAGAATTTGGGTATATAAAGTAGATACAAACACTTTCAGATTCTTATAAT 908
 601 ATCGGAGGAATCAAGAATTTGGGTATATAAAGTAGATACAAACACTTTCAGATTCTTATAAT 660
 909 TACAGTGGAAACAGAAAAATTTAAAGATAAAGATCTTTTAACTCAGTTAGATCAATTTGTT 968
 661 TACAGTGGAAACAGAAAAATTTAAAGATAAAGATCTTTTAACTCAGTTAGATCAATTTGTT 720
 969 GATTTTAACTGTCATCTGCTTGTGATCGACAAAGTTCCTTCAATGTTTTCATGCCAAGAC 1028
 721 GATTTTAACTGTCATCTGCTTGTGATCGACAAAGTTCCTTCAATGTTTTCATGCCAAGAC 780
 1029 AAGCTACAAACAGAGCCAGCCATGTGGATTACTTAAAGATGTTGGCTTAGTAAAGAG 1088
 781 AAGCTACAAACAGAGCCAGCCATGTGGATTACTTAAAGATGTTGGCTTAGTAAAGAG 840
 1089 GAAGTAGATGTGGCAGTCTAATCTGCCCGCAGATGTTTAAAGAGAGAGGGCAAGCAAGT 1148
 841 GAAGTAGATGTGGCAGTCTAATCTGCCCGCAGATGTTTAAAGAGAGAGGGCAAGCAAGT 900
 1149 GCTTTGACCTCGAGCCCTCCGAAAAATGAAGATTTATCTCTTAAATGATTCAAAATCAAGA 1208
 901 GCTTTGACCTCGAGCCCTCCGAAAAATGAAGATTTATGCTTAAATGATTCAAAATCAAGA 960
 1209 GATGAAAAATTTCAATTAACCTGACCTTTTCCCTTCAGGAGATGAGACTGTTTAAACAA 1268
 961 GATGAAAAATTTCAATTAACCTGACCTTTTCCCTTCAGGAGATGAGACTGTTTAAACAA 1020
 1269 TCTGCACAAGAGACTCAAAAAGTTTACACCTTAAAGGATAATGATGTAATCCAAAGATTCC 1328
 1021 TCTGCACAAGAGACTCAAAAAGTTTACACCTTAAAGGATAATGATGTAATCCAAAGATTCC 1080
 1329 TCTCAGCTTTACATGTTTCCAGTAAAGATGTGCGCTCCTCATTTGCTCTGCTTCTCTCGG 1388
 1081 TCTCAGCTTTACATGTTTCCAGTAAAGATGTGCGCTCCTCATTTGCTCTGCTTCTCTCGG 1140
 1389 TCTGCGTCTGTGTGGATCAATTAATGAAAAGTAAAGCACCGGGTGAATTTTACCTCAG 1448
 1141 TCTGCGTCTGTGTGGATCAATTAATGAAAAGTAAAGCACCGGGTGAATTTTACCTCAG 1200
 1449 CATGAACATAAAGATAATATCAAGATGCAGTGCATATACATGAAGAATAACAGAACAGT 1508
 1201 CATGAACATAAAGATAATATCAAGATGCAGTGCATATACATGAAGAATAACAGAACAGT 1260

QY 1509 GTTGTCTAGGTGGGAAACCAATTCAAGAGAAATGATCTTTTGAACAGGAAATATAA 1568
Db 1261 GTTGTCTAGGTGGGAAACCAATTCAAGAGAAATGATCTTTTGAACAGGAAATATAA 1320
QY 1569 AGCATACTCTTCAGTCATTAAATGAGGATGGAAGAGAGATAGATCTGACACAG 1628
Db 1321 AGCATACTCTTCAGTCATTAAATGAGGATGGAAGAGAGATAGATCTGACACAG 1380
QY 1629 ACAGTAATCAGAGCTGAGTCTTTGGATGGTGGTGACACCAAGTCTACAGTTGAGAACT 1688
Db 1381 ACAGTAATCAGAGCTGAGTCTTTGGATGGTGGTGACACCAAGTCTACAGTTGAGAACT 1440
QY 1689 CAAGAGGGGCTTTCTGCACTCATGTCCTCAGAGTCTTCGATTTGAGAGGTTTAT 1748
Db 1441 CAAGAGGGGCTTTCTGCACTCATGTCCTCAGAGTCTTCGATTTGAGAGGTTTAT 1500
QY 1749 AATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGATGAA 1808
Db 1501 AATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTAATTAATTTGATGAA 1560
QY 1809 GGGCAAAAGGTGCCCACTAAATTAATGATGCTGAACCTTGATGCCCTTTCTGACAGAACAG 1868
Db 1561 GGGCAAAAGGTGCCCACTAAATTAATGATGCTGAACCTTGATGCCCTTTCTGACAGAACAG 1620
QY 1869 TATCTTCAGACCACTACATAAAGTCTTTTGAAGAAATGTAATGACTCTAAATGCAAA 1928
Db 1621 TATCTTCAGACCACTACATAAAGTCTTTTGAAGAAATGTAATGACTCTAAATGCAAA 1680
QY 1929 ATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATTTCAAT 1988
Db 1681 ATGAATCAGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATTTCAAT 1740
QY 1989 GCAGAACAGAGCTATTTGGGAAGTCATGGTATTAATTAATTTGGAAGACAGTTGAT 2048
Db 1741 GCAGAACAGAGCTATTTGGGAAGTCATGGTATTAATTAATTTGGAAGACAGTTGAT 1800
QY 2049 AACAATAACAATAGAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGTTCAA 2108
Db 1801 AACAATAACAATAGAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGTTCAA 1860
QY 2109 CAAGGGTTACCTACAGTAAGTCTGAGATTAACAATCAATATCAGTCTCTGATATTAAC 2168
Db 1861 CAAGGGTTACCTACAGTAAGTCTGAGATTAACAATCAATATCAGTCTCTGATATTAAC 1920
QY 2169 AGTCAATCTGTGGAGGGCCAGACCTAAGCAATTTGTTAGCCCTTCCATCAAGAACAGG 2228
Db 1921 AGTCAATCTGTGGAGGGCCAGACCTAAGCAATTTGTTAGCCCTTCCATCAAGAACAGG 1980
QY 2229 AGTTCAAGGACCTGAATAGCCAGATGTTCCAGATACAATAGAAAGTGAACCCAGACA 2288
Db 1981 AGTTCAAGGACCTGAATAGCCAGATGTTCCAGATACAATAGAAAGTGAACCCAGACA 2040
QY 2289 GCAGATACCGTTGTTCCAACTACTGTGCTATGATTTCTACAGCTGATCCACAGGTTAGC 2348
Db 2041 GCAGATACCGTTGTTCCAACTACTGTGCTATGATTTCTACAGCTGATCCACAGGTTAGC 2100
QY 2349 TTCAACTTAATTAATGATATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGTAAC 2408
Db 2101 TTCAACTTAATTAATGATATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGTAAC 2160
QY 2409 GCAATAGAGATTCGTGACTGGAACAACATTTGGAAGAGGCTTGGTTTGGGCCAGAAA 2468
Db 2161 GCAATAGAGATTCGTGACTGGAACAACATTTGGAAGAGGCTTGGTTTGGGCCAGAAA 2220
QY 2469 CAGCCTACTCGGTTTCTGTATTCAGAGCTCCAACTGTATGAATCTGCCAAGTCAAAATTT 2528
Db 2221 CAGCCTACTCGGTTTCTGTATTCAGAGCTCCAACTGTATGAATCTGCCAAGTCAAAATTT 2280
QY 2529 ACTTTTACCAACCGCGACACCAATTCGCGAGCATGTGGGAAAGTATTTGTGGTGTCTGT 2588
Db 2281 ACTTTTACCAACCGCGACACCAATTCGCGAGCATGTGGGAAAGTATTTGTGGTGTCTGT 2340

QY 2589 TGTATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATATGTGTAGTCTGC 2648
Db 2341 TGTATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATATGTGTAGTCTGC 2400
QY 2649 TATGAAAATATTAGTAAAGCTCAGSCAATTTGAAAGGATGATGAGTCCAACCTGGTTCTAAT 2708
Db 2401 TATGAAAATATTAGTAAAGCTCAGSCAATTTGAAAGGATGATGAGTCCAACCTGGTTCTAAT 2460
QY 2709 CTTAAGTCTAATCATTTCTGATGAATGTACTACTGTCAGCCCTCTCCAGGAACCAACA 2768
Db 2461 CTTAAGTCTAATCATTTCTGATGAATGTACTACTGTCAGCCCTCTCCAGGAACCAACA 2520
QY 2769 TCCAGTATACCTTCCACAGCAACTTTTGCAGTCTCAGCACTTAAACCAACCCAGGTTTGA 2828
Db 2521 TCCAGTATACCTTCCACAGCAACTTTTGCAGTCTCAGCACTTAAACCAACCCAGGTTTGA 2580
QY 2829 GGAATGTTTCCAAAGAAACAGAGAGATGTTGTTTGCAGATGGTATATTTGCCCAATGCT 2888
Db 2581 GGAATGTTTCCAAAGAAACAGAGAGATGTTGTTTGCAGATGGTATATTTGCCCAATGCT 2640
QY 2889 GGAATGTTGAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTCTGGAAGCTTTAGT 2948
Db 2641 GGAATGTTGAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTCTGGAAGCTTTAGT 2700
QY 2949 CCTCTCTCACCTGATGTGCTATGACAGTAAACACAGTGGATCATTCOCATTTCTACTACA 3008
Db 2701 CCTCTCTCACCTGATGTGCTATGACAGTAAACACAGTGGATCATTCOCATTTCTACTACA 2760
QY 3009 GTGGAAGGCCAAAACAATGACAGAGATATTAAGAAGATGAGATTAATTCAGAGTCTCT 3068
Db 2761 GTGGAAGGCCAAAACAATGACAGAGATATTAAGAAGATGAGATTAATTCAGAGTCTCT 2820
QY 3069 ATTTCTCAGGTTCCATCAGTGGAAAAATTTGCTATGAACACAGGAAATGAGGGGTACCT 3128
Db 2821 ATTTCTCAGGTTCCATCAGTGGAAAAATTTGCTATGAACACAGGAAATGAGGGGTACCT 2880
QY 3129 ACTCTGGTTCAATTTACTAGATGATGATGTTTTCAGAAACTGAGAAACCATCTAGT 3188
Db 2881 ACTCTGGTTCAATTTACTAGATGATGATGTTTTCAGAAACTGAGAAACCATCTAGT 2940
QY 3189 CCTACTGGTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATATATAGGTTA 3248
Db 2941 CCTACTGGTGTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATATATAGGTTA 3000
QY 3249 CTGTGTATTTAAACAAGTATGCTGCAATAAGATGATGTTCTACCTAATGATGAGGAC 3308
Db 3001 CTGTGTATTTAAACAAGTATGCTGCAATAAGATGATGTTCTACCTAATGATGAGGAC 3060
QY 3309 AGTTTCCCCCACTTCTGGTTCATCTGGAGAAAGGATCAGTGCCTGCTAGTAGAAGAA 3368
Db 3061 AGTTTCCCCCACTTCTGGTTCATCTGGAGAAAGGATCAGTGCCTGCTAGTAGAAGAA 3120
QY 3369 CATCCATCTCATGAGCAGATCAATTTTGTCTTTGAAGGTGAAGGCTTTTCATCTGTACA 3428
Db 3121 CATCCATCTCATGAGCAGATCAATTTTGTCTTTGAAGGTGAAGGCTTTTCATCTGTACA 3180
QY 3429 TTTGTCTTAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATCTCTCAGACAA 3488
Db 3181 TTTGTCTTAATGCTTAATCTACTCGTGAATGTCAAATTCATATTTTATCTCTCAGACAA 3240
QY 3489 TATTGTGTCTTTTCAACCAATGGATTGCGCTGGGACAGGCAAGAAATTTATTTCTA 3548
Db 3241 TATTGTGTCTTTTCAACCAATGGATTGCGCTGGGACAGGCAAGAAATTTATTTCTA 3300
QY 3549 TTGTTAAGTTTGCAAAATGAAGATATTTCTTCAAGGACATCTTCAGACTATTTATCAC 3608
Db 3301 TTGTTAAGTTTGCAAAATGAAGATATTTCTTCAAGGACATCTTCAGACTATTTATCAC 3360
QY 3609 ATATATAAGGATGCTCTAAAGGAAATACATAGAAAACTTCGACAAATATTACCTTTACT 3668
Db 3361 ATATATAAGGATGCTCTAAAGGAAATACATAGAAAACTTCGACAAATATTACCTTTACT 3420
QY 3669 GAGAGTTTCTCAGTAGCAAGGATCAGGAGGATTTCTCTTTTATTATACACTTTTTCAG 3728

Db 3421 GAGAGTTTCTCAGTAGCAAGGATCACGAGGATTCCTGTTTATTACACCTACTTTTCAG 3480
 QY 3729 AAACCTTGATGCTCTCATACCAAGTAATCTTTCTTCTGGAATTTCTTATCCAGAAG 3788
 Db 3481 AAACCTTGATGCTCTCATACCAAGTAATCTTTCTTCTGGAATTTCTTATCCAGAAG 3540
 QY 3789 CTTGAGATTCCTGGGCAAGGTTTCTTCTATGCGTTTAAATGTTGAGATGGGTGCAGAA 3848
 Db 3541 CTTGAGATTCCTGGGCAAGGTTTCTTCTATGCGTTTAAATGTTGAGATGGGTGCAGAA 3600
 QY 3849 TATAAGCATATCTCTCTCTACACAGCATCAGAGCCGAAACCTCTTTTGGAGAA 3908
 Db 3601 TATAAGCATATCTCTCTCTACACAGCATCAGAGCCGAAACCTCTTTTGGAGAA 3660
 QY 3909 ATAGGACACATATATGAACCTTACTTGTGACCTTGGAAATACCAGTATACCTTGCAT 3968
 Db 3661 ATAGGACACATATATGAACCTTACTTGTGACCTTGGAAATACCAGTATACCTTGCAT 3720
 QY 3969 AATATAGATCACTGTTGATTCATATGGAATGGAAAGCTGCATATAAAATACCAGG 4028
 Db 3721 AATATAGATCACTGTTGATTCATATGGAATGGAAAGCTGCATATAAAATACCAGG 3780
 QY 4029 AAAAAGTACAGTGATGAATGAAGTACTAAATCTTCCATGAGCATGTCATTAGCAT 4088
 Db 3781 AAAAAGTACAGTGATGAATGAAGTACTAAATCTTCCATGAGCATGTCATTAGCAT 3840
 QY 4089 GGAGCAATTTACGACAGAGAGATTCCTCATCTAGTCTGTATACAGATGATGAATT 4148
 Db 3841 GGAGCAATTTACGACAGAGAGATTCCTCATCTAGTCTGTATACAGATGATGAATT 3900
 QY 4149 TATGAACACAGGCAACAGTGCACCTGCGCATCTCAGAAAGTACAGGTGCAAGTTT 4208
 Db 3901 TATGAACACAGGCAACAGTGCACCTGCGCATCTCAGAAAGTACAGGTGCAAGTTT 3860
 QY 4209 GTGGTATTCAATGGAGCTCTAAAAACATCTTCAGGATTTCTGTAAGTCCAGCATAGTT 4268
 Db 3961 GTGGTATTCAATGGAGCTCTAAAAACATCTTCAGGATTTCTGTAAGTCCAGCATAGTT 4020
 QY 4269 GAAGATGCTTTAATGGTACAAATTAATCCAGAGACCATGAATGCTTCGGCTAGCTTTA 4328
 Db 4021 GAAGATGCTTTAATGGTACAAATTAATCCAGAGACCATGAATGCTTCGGCTAGCTTTA 4080
 QY 4329 CGAGAACAGAAAGACTTTAAATATACATGCGGAAAGTTGATGAGTACAGCTCAGAGAA 4388
 Db 4081 CGAGAACAGAAAGACTTTAAATATACATGCGGAAAGTTGATGAGTACAGCTCAGAGAA 4140
 QY 4389 TACGTGGATATCTGCTGGTATGATCTGAAGAAAGAAACAAAGAGTTATCAGTTCA 4448
 Db 4141 TACGTGGATATCTGCTGGTATGATCTGAAGAAAGAAACAAAGAGTTATCAGTTCA 4200
 QY 4449 GTGATGAATATCATTAACAGGATTTTCAAGTCAAAAAATAAACTGGAAGCAGATTTT 4508
 Db 4201 GTGATGAATATCATTAACAGGATTTTCAAGTCAAAAAATAAACTGGAAGCAGATTTT 4260
 QY 4509 GAAACCGATGAGAGATTTGTAATGTACCGAGTGTCTTACTTTCTTAAAGGACCAAGAT 4568
 Db 4261 GAAACCGATGAGAGATTTGTAATGTACCGAGTGTCTTACTTTCTTAAAGGACCAAGAT 4320
 QY 4569 TTATCTATTTTATCACTCTTATCATCTTTCATGTTTCAAGAAATAGCATGGCTGTAGTGCT 4628
 Db 4321 TTATCTATTTTATCACTCTTATCATCTTTCATGTTTCAAGAAATAGCATGGCTGTAGTGCT 4380
 QY 4629 CGCTGTGCTCCCTCACTGAAACCTTAAAAAGTAAATGGATGATAAAATGGACTCAGA 4688
 Db 4381 CGCTGTGCTCCCTCACTGAAACCTTAAAAAGTAAATGGATGATAAAATGGACTCAGA 4440
 QY 4689 GTTTCATGTACATGATATGGTTGAATTTTACGACAGATCTGAGGCCAATCTTCGCT 4748
 Db 4441 GTTTCATGTACATGATATGGTTGAATTTTACGACAGATCTGAGGCCAATCTTCGCT 4500
 QY 4749 CAGCATTTCTAAATGATCTTGTAGTGTCTCTGATCTGTATCCATGCTGGGACCTCC 4808

Db 4501 CAGCATTTCTAAATGATCTTGATAGTCTCTGTGATACCTGTGATCCATCGTGGACCTCC 4560
 QY 4809 AACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTTCATTATAGAACATCTTTTATAG 4868
 Db 4561 AACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTTCATTATAGAACATCTTTTATAG 4620

RESULT 5
 BC032227
 LOCUS
 DEFINITION
 Homo sapiens zinc finger, FYVE domain containing 16, mRNA (CDNA
 clone MGC:34354 IMAGE:5204515), complete cds.
 ACCESSION
 BC032227
 VERSION
 BC032227.1
 KEYWORDS
 GI:21619001
 SOURCE
 MGC.
 ORGANISM
 Homo sapiens (human)
 Eukaryota
 Chordata
 Craniata
 Vertebrata
 Euteleostomi
 Mammalia
 Eutheria
 Primates
 Catarrhini
 Hominoidea
 Homo
 1 (bases 1 to 4499)
 REFERENCE
 1
 AUTHORS
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.B., Brownstein,M.J., Uscin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.C., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahy,J., Hellon,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shetchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickinson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,U., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
 Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 4499)
 REFERENCE
 2
 AUTHORS
 Strausberg,R.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nih.gov
 Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,
 Antier,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Dietrich,N.L., Grant,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
 Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantripo,S., Thomas,P.J., Touchman,J.W.,
 Tsurguen,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 49 Row: f Column: 4

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7662047.

FEATURES	Source	Location/Qualifiers		QY	Db
		1.. 4499			
	gene	/organism="Homo sapiens"		365	CCACTGCTCAGTCTTCTTCAGAGTTGGCTTCCTCAGCGGAACTTCATTCCTCCCAAGA 424
		/mol_type="mRNA"		255	CCACTGCTCAGTCTTCTTCAGAGTTGGCTTCCTCAGCGGAACTTCATTCCTCCCAAGA 314
		/db_xref="taxon:9606"		425	CCAAGATGCGGTTAATAGTTGCTCATCAGAAACAAGCTATGGAACAATGAGAGTTC 484
		/clone="MGC:34354 IMAGE:5204515"		315	CCAAGATGCGGTTAATAGTTGCTCATCAGAAACAAGCTATGGAACAATGAGAGTTC 374
		/tissue types="Lung, Spleen, fetal, pooled"		485	CTGTAATGAAAAACACCTCAAGGAGCTTACTCTATACAAATGAAAAAATGTAACAGG 544
		/clone lib="NIH MGC_122"		375	CCTGAATGAAAAACACCTCAAGGAGCTTACTCTATACAAATGAAAAAATGTAACAGG 434
		/lab hosts="DH10B"		545	ACTTGATCTTCTTCTTCTGATGGTGGTACTTCAGATGAAATCCAGCCGTTATATAT 604
		/notes="vector: pCMV-SPORT6"		435	ACTTGATCTTCTTCTTCTGATGGTGGTACTTCAGATGAAATCCAGCCGTTATATAT 494
		1.. 4499		605	GGGACGATGATGATAAACCCTATCTGTGATCTGATAGTACATGGTAACTTAGTTCATGC 664
		/genes="ZFVVE16"		495	GGGACGATGATGATAAACCCTATCTGTGATCTGATAGTACATGGTAACTTAGTTCATGC 554
	CDS	/notes="synonyms: KIAA0305, ENDOFIN"		665	ACCATAAGTGAAGAAGATATTTAAAAAATTTGCGAGATGATTTTAAAGTCTAAATGCAGA 724
		/db_xref="LocusID:9765"		555	ACCATAAGTGAAGAAGATATTTAAAAAATTTGCGAGATGATTTTAAAGTCTAAATGCAGA 614
		139.. 3756		725	TTCTTTGATTGGATTGGATTATCTTCAGTGTGATGATCTCCCTGTGTTCTTCAACAGA 784
		/product="ZFVVE16 protein"		615	TTCTTTGATTGGATTGGATTATCTTCAGTGTGATGATCTCCCTGTGTTCTTCAACAGA 674
		/protein_id="AAH32227.1"		785	CCATGATGATGATCTGTCAGAGAACACAGATGATGATGATGATGATGATGATGATGATGAT 844
		/db_xref="GI:21619002"		675	CCATGATGATGATCTGTCAGAGAACACAGATGATGATGATGATGATGATGATGATGATGAT 734
		/translat="MDSYFAAASVSDLDKLLDFFQNPDEQYLDQVQAYDSNHCBSV		845	AGAAATCGGAGGAATCAAAAGATTTGGGTATATAAAGTAGATACAACTTTCAGATTCTCTA 904
		SESSORTSLLPKQDVNSVSCASSETSYGNSNNEKTLGLTIGNEKNVTGLDL		735	AGAAATCGGAGGAATCAAAAGATTTGGGTATATAAAGTAGATACAACTTTCAGATTCTCTA 794
		LIGLDSVSDTFCVSDTSDHSDTVEQNDTSSELRNREIGGKELGKIVDTTLSDS		905	TAATACAGTGAAGACAGAAATTTAAAGATATAAAGATCTTTAATCAGTGTAGATCAAT 964
		INVSTENKDKKIFNOLSVIPNMSALTROSSKMFHAKDLQHKOPCGLLKDVG		795	TAATACAGTGAAGACAGAAATTTAAAGATATAAAGATCTTTAATCAGTGTAGATCAAT 854
	misc_feature	LVXEEVAVITAECLKRGKTSALTSLPNCEDLCLNDSNRDENFKLPDRSFRD		965	TGTTGATTTTAACTGTCATCTGCTTGTGCTGACGAAAGTTCCTGCTGCTGCTGCTGCTGCTGCT 1024
		KTWIKQAEQSDSLDKNDVTDSSSALHVSVDVPSLSCLPASGSMGSLISLK		855	TGTTGATTTTAACTGTCATCTGCTTGTGCTGACGAAAGTTCCTGCTGCTGCTGCTGCTGCTGCT 914
		ARGDFLPQHEKNIQDAVTHIBIQNSVVLGEPKENDLLKOEKCKSLLOSLEG		1025	AGACAAAGCTACACACAGAGCGCCATGTGGATTAATAAAGATGTTGGCTTAGTAA 1084
		MDKIDPDTVRAESLDGDTSTVVEQEGLSGTHVPESSDCEGFTTFSSNDM		915	AGACAAAGCTACACACAGAGCGCCATGTGGATTAATAAAGATGTTGGCTTAGTAA 974
		DGDLDYENIDEGAGSLISDAFLTEOYLQTNIKSPENVNDSKSNQNDM		1085	AGAGGATGATGATGTCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1144
		KGLDDGNNIYNBAAGAGESHGNIICETVDKONTIENGSLGKSTIPVQGLP		975	AGAGGATGATGATGTCGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1034
		TSKEITNLSVDINSQSVNSGYPKQFSLPRTSSKDLNKPDPDPTIESPSTAD		1145	AACTGCTTTGACCTGCGAGCTTCGGAATAATGAAGATTTATGCTTAAATGATTCAAAATTC 1204
		QVTWVPSAPNCNQCQVFTKRRHRCACQKFCVGCNCRKXLYLEKEARVCV		1035	AACTGCTTTGACCTGCGAGCTTCGGAATAATGAAGATTTATGCTTAAATGATTCAAAATTC 1094
		PGVGLCSKQKRYFADGILPNCVADTTLKLSGKRCSEDFSPSPDVPMTVNDV		1205	AAAGATGATAAATTCAGATCTTCTTCCTTCAGGAGATGAGTGTATGATGATGATGATGATGATGATGAT 1264
		HSHSTVKEKNNEGDIITREIIOQSPVSVSEKLSMNTGNEGLPTSGSFTLDDVDF		1095	AAAGATGATAAATTCAGATCTTCTTCCTTCAGGAGATGAGTGTATGATGATGATGATGATGATGATGAT 1154
	ORIGIN	AKTBEPSPTGLVNSNLPASISDYRLDINKYCNKISLAPNEDSLPPLLVASG		1265	ACAATCTGCAACAAGAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGATGATGATGATGATGATGAT 1324
		EKGSPVVEBHEHQIILLDEGGFPFTFVLNANLVNKIIFVSSDKYWFSTNG		1155	ACAATCTGCAACAAGAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGATGATGATGATGATGATGAT 1214
		LHGLQAEIILLCLNEDPTPKDIFPLTIYDALKRGIENDNITFTSFJSS		1325	TTCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTGCCGTCCCTCATTTGCTGCTGCTTCC 1384
		KHGGFLFITPTFKQLDLSLPSNPFLLGIQKLEIPWAKVPMRLRLGAEYKXK		1215	TTCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTGCCGTCCCTCATTTGCTGCTGCTTCC 1274
		F"		1385	TGGCTCTGGGCTCTATGTGTGGATCATTATTTGAAGTAAAGACCGGGGTGATTTTTTACC 1444
		2362.. 2556		1275	TGGCTCTGGGCTCTATGTGTGGATCATTATTTGAAGTAAAGACCGGGGTGATTTTTTACC 1334
		/note="FYVE, Region: FYVE zinc finger. The FYVE zinc finger is named after four proteins that it has been found in: Fabi, YOTB/ZK632.12, Vac1, and BEAL. The FYVE finger has been shown to bind two Zn++ ions. The FYVE finger has eight potential zinc coordinating cysteine positions. Many members of this family also include two histidines in a motif R/HHC+XCG, where + represents a charged residue and X any residue. We have included members which do not conserve these histidine residues but are clearly related"		1445	TCAGCATGAACATAAAGATATAACAGATGCGAGTGTATACATGAAGAAATACAGAA 1504
		/db_xref="CDD:pfam01363"			
		Query Match 54.9%; Score 3643.8; DB 9; Length 4499;			
		Best Local Similarity 99.5%; Pred No. 0;			
		Matches 3654; Conservative 0; Mismatches 17; Indels 0; Gaps 0;			
		185 ATATCTCTCCAGGATGTTCTCAAGGCATACAAGATTAATTTCTGAATAAGTCTGAGG 244			
		75 AAAGCTCCGACGGGCTGTAGGAGGCATACAAGAAATTAATTTCTGAATAAGTCTGAGG 134			
		245 TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCACTTGACAACTCCCTTTGATGATT 304			
		135 TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCACTTGACAACTCCCTTTGATGATT 194			
		305 TGAACAGAACCCAGATGAACAGATTATCTCGCAGATGTACAAATGCATATGATTTCTAA 364			
		195 TGAACAGAACCCAGATGAACAGATTATCTCCAGATGTACAAATGCATATGATTTCTAA 254			

1335 TCAGCATGACATAAAGATAATATACAGATGAGTACATATACATGAGAAATACAGAA 1394
1505 CAGTGTGTGTAGTGGGAAACCAATTCAGAGAGAAATGATCTTTTGAACACAGGAAAAATG 1564
1395 CAGTGTGTGTAGTGGGAAACCAATTCAGAGAGAAATGATCTTTTGAACACAGGAAATG 1454
1565 TAAAGCATACCTCTCAGTGTCTTAAATTCAGAGGATGGAAGACAGAAAGATAGATCTCTGA 1624
1455 TAAAGCATACCTCTCAGTGTCTTAAATTCAGAGGATGGAAGACAGAAAGATAGATCTCTGA 1514
1625 CCAGACAGTAATCAGACGTGAGTCTTTGGATGGTGGTGCACACCACTTCTACAGTTGTAGA 1684
1515 CCAGACAGTAATCAGACGTGAGTCTTTGGATGGTGGTGCACACCACTTCTACAGTTGTAGA 1574
1685 ATCTCAAGAGGGCTTTCTGGCACTCATGTCCAGAGCTTTCTGATTTGTGAGGTTT 1744
1575 ATCTCAAGAGGGCTTTCTGGCACTCATGTCCAGAGCTTTCTGATTTGTGAGGTTT 1634
1745 TATTAACTCTTTTCAAGCAATGATATGATGGGCAAGCTTAGATTTACTTTTATATGA 1804
1635 TATTAACTCTTTTCAAGCAATGATATGATGGGCAAGCTTAGATTTACTTTTATATGA 1694
1805 TGAAGGCGCAAAAGTGGCCCACTAAATAGTATGCTGAACTTTGATGCTTTCTGACAGA 1864
1695 TGAAGGCGCAAAAGTGGCCCACTAAATAGTATGCTGAACTTTGATGCTTTCTGACAGA 1754
1865 ACAGTATCTTCAGACCACTACATTAAGTCTTTTGAAGAAATGTAATGATCTTAATC 1924
1755 ACAGTATCTTCAGACCACTACATTAAGTCTTTTGAAGAAATGTAATGATCTTAATC 1814
1925 GCAATGAATCAGATAGATATGAAAGGCTTAGATGCAAAACATCAATATATATATTT 1984
1815 GCAATGAATCAGATAGATATGAAAGGCTTAGATGCAAAACATCAATATATATATTT 1874
1985 CAATCAGAGCAGGAGCTATTTGGGAAAGTCATGGTATTAATATATATTTTGTGAACAGT 2044
1875 CAATCAGAGCAGGAGCTATTTGGGAAAGTCATGGTATTAATATATATTTTGTGAACAGT 1934
2045 TGATAACAAATAACAATAGAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGT 2104
1935 TGATAACAAATAACAATAGAAATGGCTTTCTTTAGGAGAAAGCACTATTTCCAGT 1994
2105 TCAACAGGGTTACCTACCACTAGTCTGAGATTAACAAATCAATATATATATATATAT 2164
1995 TCAACAGGGTTACCTACCACTAGTCTGAGATTAACAAATCAATATATATATATATAT 2054
2165 TAACAGTCAATCTGTGGAGGGCCAGACCTTAAGCAATTTGTAGCTTCCATCAAGAAC 2224
2055 TAACAGTCAATCTGTGTGGAGGGCCAGACCTTAAGCAATTTGTAGCTTCCATCAAGAAC 2114
2225 AAGGAGTTCAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAGTGAACCCAG 2284
2115 AAGGAGTTCAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAGTGAACCCAG 2174
2285 CACAGCAGATACGTTGTTCCAACTCATCTGTGCTATAGATTTACAGCTGATCCACAGGT 2344
2175 CACAGCAGATACGTTGTTCCAACTCATCTGTGCTATAGATTTACAGCTGATCCACAGGT 2234
2345 TAGCTTCAACTCAATATACATTGATATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGT 2404
2235 TAGCTTCAACTCAATATACATTGATATAGAAAGTAAATTTCTGAAGGTGGATCTAGTTTCGT 2294
2405 AACTGCAAAATGAAGATCTGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2464
2295 AACTGCAAAATGAAGATCTGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 2354
2465 GAAACAGCCTACTTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2524
2355 GAAACAGCCTACTTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2414
2525 ATTTACTTTTACAAAACGGGACACCAATGCGAGCATGCGGAAAGTATTTTGTGGTGT 2584

2415 ATTTACTTTTACAAAACGGGACACCAATGCGAGCATGCGGAAAGTATTTTGTGGTGT 2474
2585 CTGTGTGTATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGATATGTGTACT 2644
2475 CTGTGTGTATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGATATGTGTACT 2534
2645 CTGTGTGTATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGATATGTGTACT 2704
2535 CTGTGTGTATAGGAAGTGTAAACTGCAATATCTAGAAAAAGGAGCAAGATATGTGTACT 2594
2705 TAATCTTAAGTCTAAATCAATCTGATGAATGTACTACTGCTCCAGCTTCCCTCAGGAGAACCA 2764
2595 TAATCTTAAGTCTAAATCAATCTGATGAATGTACTACTGCTCCAGCTTCCCTCAGGAGAACCA 2654
2765 AACATCCAGTATACCTTCCAGCAGCAACTTTTGCAGCTTCCAGCACTTAAACACCAAGGTGT 2824
2655 AACATCCAGTATACCTTCCAGCAGCAACTTTTGCAGCTTCCAGCACTTAAACACCAAGGTGT 2714
2825 TGAAGCACTATGTTCCAAAGAAACAGAGAGATGATGTTTGCAGATGATATATGCCCCAA 2884
2715 TGAAGCACTATGTTCCAAAGAAACAGAGAGATGATGTTTGCAGATGATATATGCCCCAA 2774
2885 TGGTGAAGTTGCAGATACAAACAAAATTTATCATCTGGAAGTAAAGATGTTCTGAAGACTT 2944
2775 TGGTGAAGTTGCAGATACAAACAAAATTTATCATCTGGAAGTAAAGATGTTCTGAAGACTT 2834
2945 TAGTCTCTCTCAGCTGATGCTGCTATGACAGTAAACACAGTGGATCATTCCTATCTAC 3004
2835 TAGTCTCTCTCAGCTGATGCTGCTATGACAGTAAACACAGTGGATCATTCCTATCTAC 2894
3005 TACAGTGGAAAGCCAAACAAATGACAGAGATATTAACAGAAATGAGATAATTCAGAG 3064
2895 TACAGTGGAAAGCCAAACAAATGACAGAGATATTAACAGAAATGAGATAATTCAGAG 2954
3065 TCTATTTCTCAGGTTCCATCAGTGGAAATAATGCTATGAAACACAGAAATAGGGGTT 3124
2955 TCTATTTCTCAGGTTCCATCAGTGGAAATAATGCTATGAAACACAGAAATAGGGGTT 3014
3125 ACTTCTCTCTGTTTCAATTTACACTAGATGATGATGTTTTCAGAGAACTGAGAACCATC 3184
3015 ACTTCTCTCTGTTTCAATTTACACTAGATGATGATGTTTTCAGAGAACTGAGAACCATC 3074
3185 TAGTCTCTCTGTTTCAATTTACACTAGATGATGATGTTTTCAGATTTTTCAGATTTAG 3244
3075 TAGTCTCTCTGTTTCAATTTACACTAGATGATGATGTTTTCAGATTTTTCAGATTTAG 3134
3245 GTTACTGTGTATATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3304
3135 GTTACTGTGTATATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3194
3305 GGACAGTTTGGCCCACTTCTGTTTGCATCTGGAGAAAGGATCAGTGCCTGTAGTAGA 3364
3195 GGACAGTTTGGCCCACTTCTGTTTGCATCTGGAGAAAGGATCAGTGCCTGTAGTAGA 3254
3365 AGAACATCCATCTCATGAGCAGATCATTTTGTCTTTCAGAGGTGAGGCTTTCATCTGT 3424
3255 AGAACATCCATCTCATGAGCAGATCATTTTGTCTTTCAGAGGTGAGGCTTTCATCTGT 3314
3425 TACATTTGTCTTAAATGCTAATCTACTGCTGAATGCTCAATTTTATTTTATTTCTCAGA 3484
3315 TACATTTGTCTTAAATGCTAATCTACTGCTGAATGCTCAATTTTATTTTATTTCTCAGA 3374
3485 CAAATATTTGGTACTTTTCAACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 3544
3375 CAAATATTTGGTACTTTTCAACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 3434
3545 TCTATTTGTATGTTTGGCCAAATGAGATACATTTTCCCTTTCAGACATTTTATTTAT 3604
3435 TCTATTTGTATGTTTGGCCAAATGAGATACATTTTCCCTTTCAGACATTTTATTTAT 3494
3605 CACCATATATAGGATGCTTAAAGGAAATAATATAGAAACTTTGGAACAATATTTACCTT 3664
3495 CACCATATATAGGATGCTTAAAGGAAATAATATAGAAACTTTGGAACAATATTTACCTT 3554

Qy	3665	TACTGAGAGTTTCTCAGTAGCAGGATCAAGGAGTTCCTGTTTATTACACCTACTTT	3724
Db	3555	TACTGAGAGTTTCTCAGTAGCAGGATCAAGGAGTTCCTGTTTATTACACCTACTTT	3614
Qy	3725	TCAGAACTTGATGATCTCTCATTACCAAGTAATCCCTTTCTTTGTTGGAACTCTATCCA	3784
Db	3615	TCAGAACTTGATGATCTCTCATTACCAAGTAATCCCTTTCTTTGTTGGAACTCTATCCA	3674
Qy	3785	GAAGCTTGAGATCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATGGGTGC	3844
Db	3675	GAAGCTTGAGATCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATGGGTGC	3734
Qy	3845	AGATATATAG 3855	
Db	3735	AGATATATAG 3745	
RESULT 6			
AK127003			
LOCUS			
DEFINITION			
Homo sapiens cDNA FLJ45058 fis, clone BRAWH3023172, moderately similar to Mothers against decapentaplegic homolog interacting protein.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens			
Homo sapiens			
REFERENCE			
1			
AUTHORS			
Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.			
NEO human cDNA sequencing project			
Unpublished			
2 (bases 1 to 3793)			
Isogai, T. and Yamamoto, J.			
Direct Submission			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)			
NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' - end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.			
FEATURES			
source			
1. .3793			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="BRAWH3023172"			
/tissue_type="brain"			
/clone_lib="BRAWH3"			
/note="cloning vector: pME18SFL3"			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 3709; Conservative 0; Mismatches 2; Indels 259; Gaps 3;			
Query Match			
Best Local Similarity			
Matches 2663; Conservative 0; Mismatches 2; Indels 259; Gaps 3;			
Query Match			
Best Local Similarity			
Matches 70; Conservative 0; Mismatches 2; Indels 259; Gaps 3;			

Qy	2723	TTCTGATGAATGCTACTACTGTCCAGCCCTCTCAGGAGAACCAACATCCAGTATACCTTC	2782
Db	130	TTCTGATGAATGCTACTACTGTCCAGCCCTCTCAGGAGAACCAACATCCAGTATACCTTC	189
Qy	2783	ACCAGCAACTTTGCCAGTCTCAGCACTTAAACCAACCCAGGTGTTGAAGCACTATGTTCCAA	2842
Db	190	ACCAGCAACTTTGCCAGTCTCAGCACTTAAACCAACCCAGGTGTTGAAGCACTATGTTCCAA	249
Qy	2843	AGAACAGAGAGAGTATGGTTTGCAGATGGTATATGGCCATGTTGAGTGAAGTTCAGATAC	2902
Db	250	AGAACAGAGAGAGTATGGTTTGCAGATGGTATATGGCCATGTTGAGTGAAGTTCAGATAC	309
Qy	2903	AACAAATTTATCATCTCGAAGTAAAGATGTTCTCAAGACATTTAGTCTCTCTCACTCA	2962
Db	310	AACAAATTTATCATCTCGAAGTAAAGATGTTCTCAAGACATTTAGTCTCTCTCACTCA	369
Qy	2963	TGTGCTATGACAGATAAACAACAGTGATCAITCCCATTTCTACTACAGTGGAAAGCCAAA	3022
Db	370	TGTGCTAT-----	378
Qy	3023	CAATGAGACAGGAGATATTACAAGAAATGAGATAAATTCAGAGTCTCTATTCTCAGGTTC	3082
Db	379	-----	378
Qy	3083	ATCAGTGGAAAAATTTCTATGACACAGGAAATGAGGGTTTACCTACTTCTGGTTCAAT	3142
Db	379	-----	378
Qy	3143	TACACTAGATGATGATGTTTTCAGAACTGAGAACCATCTAGTCTCTACTGTTGCTCT	3202
Db	379	-----	378
Qy	3203	AGTTAACGCAATTTACCTATTGCTAGTATTTTCAGATATAGGTTACTGTGTGATATAA	3262
Db	379	-----GTATTTTCAGATATAGGTTACTGTGTGATATAA	412
Qy	3263	CAAGTATGTTCTGCAATAAGATTTAGTCTTCTACTAATGATGAGGACAGTTTCCCTCCT	3322
Db	413	CAAGTATGTTCTGCAATAAGATTTAGTCTTCTACTAATGATGAGGACAGTTTCCCTCCT	472
Qy	3323	TCGTGTTGCATCTGGAGAAAGGATCAGTGCTGTAGTAGAAGAACATCCATCTCATGA	3382
Db	473	TCGTGTTGCATCTGGAGAAAGGATCAGTGCTGTAGTAGAAGAACATCCATCTCATGA	532
Qy	3383	GCAGATCATTTTGTCTTTGAGGTTGAGGCTTTCATCTCTGTTTACATTTGTCTTAATGC	3442
Db	533	GCAGATCATTTTGTCTTTGAGGTTGAGGCTTTCATCTCTGTTTACATTTGTCTTAATGC	592
Qy	3443	TAATCTACTCTGTAATGTCATATTTTATTTCTCTCAGACAAATATTTGTTACTTTTC	3502
Db	593	TAATCTACTCTGTAATGTCATATTTTATTTCTCTCAGACAAATATTTGTTACTTTTC	652
Qy	3503	AACCAATGGAATGTCATGCTTGGGACAGGCAAGAAATTTATTTCTATTGTTTGTGCT	3562
Db	653	AACCAATGGAATGTCATGCTTGGGACAGGCAAGAAATTTATTTCTATTGTTTGTGCT	712
Qy	3563	AAATGAAGATCACTATTCCTAAGGACATCTTCAGACTATTATCACCATATAAAGATGC	3622
Db	713	AAATGAAGATCACTATTCCTAAGGACATCTTCAGACTATTATCACCATATAAAGATGC	772
Qy	3623	TCTAAAGGAAATACATAGAAACCTTGACAAATATTACCTTTACTGAGAGTTTCTCAG	3682
Db	773	TCTAAAGGAAATACATAGAAACCTTGACAAATATTACCTTTACTGAGAGTTTCTCAG	832
Qy	3683	TAGCAAGGATTCAGGAGATTCCTGTTTATTACCTACTTTTCAGAAACTTGATGATCT	3742
Db	833	TAGCAAGGATTCAGGAGATTCCTGTTTATTACCTACTTTTCAGAAACTTGATGATCT	892
Qy	3743	CTCATTTACCAAGTAATCCCTTTTCTGGAATTTCTTATCCAGAACTTTGAGATTCCTCG	3802
Db	893	CTCATTTACCAAGTAATCCCTTTTCTGGAATTTCTTATCCAGAACTTTGAGATTCCTCG	951

3803 GCACAAAGGTTTTCTATGCGTTTAAATGTTGAGATTGGTGACAGATATAAAGCATATCC 3862
Db
952 GGCACAAAGGTTTTCTATGCGTTTAAATGTTGAGATTGGTGACAGATATAAAGCATATCC 1011
Qy
3863 TGCTCCTCTAAACAAGCATCGAGCGGCGAAACCTCTTTTGGAGNAATAGGACACACTAT 3922
Db
1012 TGCTCCTCTAAACAAGCATCGAGCGGCGAAACCTCTTTTGGAGNAATAGGACACACTAT 1071
Qy
3923 TATGAACCTACTTGTGTGACCTTCGAATATACAGATATACCTTTGCATATATAGATCAACT 3982
Db
1072 TATGAACCTACTTGTGTGACCTTCGAATATACAGATATACCTTTGCATATATAGATCAACT 1131
Qy
3983 GTTGATTATATGGAATGGGNAAGCTGCATATAAATATACACGGAAAGATACAGTGA 4042
Db
1132 GTTGATTATATGGAATGGGNAAGCTGCATATAAATATACACGGAAAGATACAGTGA 1191
Qy
4043 TGTAATGAAGTACTAAATTTCTTCCAATGAGCATGTCTATAGCATTTGGAGCAAGTTTCAG 4102
Db
1192 TGTAATGAAGTACTAAATTTCTTCCAATGAGCATGTCTATAGCATTTGGAGCAAGTTTCAG 1251
Qy
4103 TACAGAACAGATTCTCATCTAGTCTGTATACAGAAATGATGGAATTTATGAAACACAGGC 4162
Db
1252 TACAGAACAGATTCTCATCTAGTCTGTATACAGAAATGATGGAATTTATGAAACACAGGC 1311
Qy
4163 CAACAGTCCCATCGCCCATCTAGAAAGTGCAGAGTGGTGCAGTTTGTGTTTCAATGG 4222
Db
1312 CAACAGTCCCATCGCCCATCTAGAAAGTGCAGAGTGGTGCAGTTTGTGTTTCAATGG 1371
Qy
4223 AGCTCTAAAAACATCTTCAGGATTTCTTGCTAAAGTCCAGCATAGTTGAAGATGCTTTAAT 4282
Db
1372 AGCTCTAAAAACATCTTCAGGATTTCTTGCTAAAGTCCAGCATAGTTGAAGATGCTTTAAT 1431
Qy
4283 GGTACAAATACCTCAGAGACCATGATGGCTTGGGCTAGCTTTACGAGAACAGAAAGA 4342
Db
1432 GGTACAAATACCTCAGAGACCATGATGGCTTGGGCTAGCTTTACGAGAACAGAAAGA 1491
Qy
4343 CTTTAAAAATTACATGTGGGAAAGTTGATGCAATGAGACCTGAGAAATACGTGGATATCTG 4402
Db
1492 CTTTAAAAATTACATGTGGGAAAGTTGATGCAATGAGACCTGAGAAATACGTGGATATCTG 1551
Qy
4403 CTGGGTAGATGCTGAAGAAAAAGGAAACAAAGAGTATACGTTTCAAGTGAATGAAATC 4462
Db
1552 CTGGGTAGATGCTGAAGAAAAAGGAAACAAAGAGTATACGTTTCAAGTGAATGAAATC 1611
Qy
4463 ATTACAGAGATTTCCAGTGAAGAAAAAGGAAACAAAGAGTATACGTTTCAAGTGAATGAA 4522
Db
1612 ATTACAGAGATTTCCAGTGAAGAAAAAGGAAACAAAGAGTATACGTTTCAAGTGAATGAA 1671
Qy
4523 GATTGTAAAAATGTACCGAGGTGTTCTACTTTCTAAAGGACAGAGATTTATCTATTTTATC 4582
Db
1672 GATTGTAAAAATGTACCGAGGTGTTCTACTTTCTAAAGGACAGAGATTTATCTATTTTATC 1731
Qy
4583 AACTTCTTATCAGTTTGCAGAAAGATAGCCATGGCTGTAGTGTCTGCTGCTGCTGCTGCT 4642
Db
1732 AACTTCTTATCAGTTTGCAGAAAGATAGCCATGGCTGTAGTGTCTGCTGCTGCTGCTGCT 1791
Qy
4643 CCTGAAAACTCTAAAAAGTATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGCACAC 4702
Db
1792 CCTGAAAACTCTAAAAAGTATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGCACAC 1851
Qy
4703 TGATATGGTTGAATTTACGAGAGATCTGAAGGCGCAACTCTGCTCAGCAATATCTAAA 4762
Db
1852 TGATATGGTTGAATTTACGAGAGATCTGAAGGCGCAACTCTGCTCAGCAATATCTAAA 1911
Qy
4763 TGATCTTGATAGTCTCTGATACCTGTGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 4822
Db
1912 TGATCTTGATAGTCTCTGATACCTGTGATCCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971
Qy
4823 ATTAGAAATAGAAATAGTGTGTTTTCATATAGAAACATCTTTTGTAGTGAAGAAATGTGCC 4882
Db
1972 ATTAGAAATAGAAATAGTGTGTTTTCATATAGAAACATCTTTTGTAGTGAAGAAATGTGCC 2031
Qy
4883 ATATTACATATTGCAACCTAATTTGTTAAAACTAACTCCAGCAGCTAAAGCTGAAATGCCA 4942

2032 ATATTACATATTGCAACCTAAATTTGTTAAAACTAACTCCAGCAGCTAAAGCTGAAATGCCA 2091
Qy
4943 CAAACACTAAAAGTATAAATATGCTGATTTTGTGAAACACATAAGCTTTGCTCTTTAGGC 5002
Db
2092 CAAACACTAAAAGTATAAATATGCTGATTTTGTGAAACACATAAGCTTTGCTCTTTAGGC 2151
Qy
5003 AGGAATGATCTTTTCAATCATTTAGCACAAATATTTAAATATCTAAAAATTTAAGAGATCC 5062
Db
2152 AGGAATGATCTTTTCAATCATTTAGCACAAATATTTAAATATCTAAAAATTTAAGAGATCC 2211
Qy
5063 ATACTTCTGCTAGCTTTCAATTAATTTAAGTACTAAAAAGACAGGATTTCTTTTAAGA 5122
Db
2212 ATACTTCTGCTAGCTTTCAATTAATTTAAGTACTAAAAAGACAGGATTTCTTTTAAGA 2271
Qy
5123 AATTATAGCATTTACTGTTTAAATGCTTAAGCCAAAGTATCTGCACCTAGGTATA 5182
Db
2272 AATTATAGCATTTACTGTTTAAATGCTTAAGCCAAAGTATCTGCACCTAGGTATA 2331
Qy
5183 CCTCTTTATGCCAAATAATGATTTTAAATGAAGGCTCTTTTCAGATGTAACCTTTAGAGGA 5242
Db
2332 CCTCTTTATGCCAAATAATGATTTTAAATGAAGGCTCTTTTCAGATGTAACCTTTAGAGGA 2391
Qy
5243 AATATCTGCTTTGTGTATATGCGAGTTAGAAATCTGTTTCTAAAGTCTGTCATAATGTA 2451
Db
2392 AATATCTGCTTTGTGTATATGCGAGTTAGAAATCTGTTTCTAAAGTCTGTCATAATGTA 2511
Qy
5303 TTTAGTGGCACAAAAACAGTTTTGAGGCTTTAGAGCTTATAATTTCTTTGAATAAAACCTG 5362
Db
2452 TTTAGTGGCACAAAAACAGTTTTGAGGCTTTAGAGCTTATAATTTCTTTGAATAAAACCTG 2511
Qy
5363 ATAACTTATTTGATTAATTTGGAGTGGAGACCTACCTCCATATTAAGATAAACTCTTTTG 5422
Db
2512 ATAACTTATTTGATTAATTTGGAGTGGAGACCTACCTCCATATTAAGATAAACTCTTTTG 2571
Qy
5423 GATTATAATCAGAAATTTTGCCCTTTTCTTCTCAAAATTTATACATATGATGTTATATAT 5482
Db
2572 GATTATAATCAGAAATTTTGCCCTTTTCTTCTCAAAATTTATACATATGATGTTATATAT 2631
Qy
5483 ATCCACATATATAGTTTTTCCCTGATTAATGGAATTAATAATTTGGGGTCTTCAGG 5542
Db
2632 ATCCACATATATAGTTTTTCCCTGATTAATGGAATTAATAATTTGGGGTCTTCAGG 2691
Qy
5543 ACTTTTGTCTCTATATTAAGTATATTTTATAGCAAGACATATTTCTGAATGTTT 5602
Db
2692 ACTTTTGTCTCTATATTAAGTATATTTTATAGCAAGACATATTTCTGAATGTTT 2751
Qy
5603 TATAAATCTTTAAATTAATGATGATTAATTTTGTATCAATGCAATTTATTTTTT 5662
Db
2752 TATAAATCTTTAAATTAATGATGATTAATTTTGTATCAATGCAATTTATTTTTT 2810
Qy
5663 TTCTCTCTCTCTCCAAACTATACCTGATTTTACCACTTCTAAAGAGTGCATGACAC 5722
Db
2811 TTCTCTCTCTCTCCAAACTATACCTGATTTTACCACTTCTAAAGAGTGCATGACAC 2870
Qy
5723 GGGCCAGATGACCCCTTGAAGTAGTCATTTATGTAGCAATAAATGAAGCCTGAAACAGGTTT 5782
Db
2871 GGGCCAGATGACCCCTTGAAGTAGTCATTTATGTAGCAATAAATGAAGCCTGAAACAGGTTT 2930
Qy
5783 TTTTACTTCCACTTTTAACTTGAAGATTTCTTGGCAACTTGGCATATTTTCAATGACAC 5842
Db
2931 TTTTACTTCCACTTTTAACTTGAAGATTTCTTGGCAACTTGGCATATTTTCAATGACAC 2990
Qy
5843 TGGTGTATAAGTATAAATTTAAATGAACATAATTTTGTATCAATGCAATTTATTTTTT 5902
Db
2991 CGGTGTAAGATATAAATTTAAATGAACATAATTTTGTATCAATTTTAAATTTCTTTATA 3050
Qy
5903 TGGTGTATTTTATTAACAGGATATTAACATAAGTAAATTCCTATGTTTGAATTTG 5962
Db
3051 TGGTGTATTTTATTAACAGGATATTAACATAAGTAAATTCCTATGTTTGAATTTG 3110
Qy
5963 TTACAGAGCTTTCCCTCTTTTACTTCAACAGCAAAAAAGTGGGGGCGCATATTTGTAGTCTG 6022

Db	3111	TTACAGAGCTTCTCTCTTACTTCAAAACAGCAAAAAGTGGGGGCATATTGTAGTCTTG	3170
Qy	6023	TCATTTAAGTTAGTAAATAATTAATCATATTATTTGATGCTTTAAACATTTCTCATCTGT	6082
Db	3171	TCATTTAAGTTAGTAAATAATTAATCATATTATTTGATGCTTTAAACATTTCTCATCTGT	3230
Qy	6083	AATATATGTTTTGATCAAAAACACTCATATATTTCAAGAAAAAGAAATATGTTAAAT	6142
Db	3231	AATATATGTTTTGATCAAAAACACTCATATATTTCAAGAAAAAGAAATATGTTAAAT	3290
Qy	6143	AGCCTCTTTTAAAGAAAAATTTATGAAGCACTCACTCACTCACTCACTCACTCACTCACT	6202
Db	3291	AGCCTCTTTTAAAGAAAAATTTATGAAGCACTCACTCACTCACTCACTCACTCACTCACT	3350
Qy	6203	TAACTCAGGATCTGAGGCTCTCAAGCTAGGAGAGACTGAGAAATTTTAAATGTTGGCAT	6262
Db	3351	TAACTCAGGATCTGAGGCTCTCAAGCTAGGAGAGACTGAGAAATTTTAAATGTTGGCAT	3410
Qy	6263	ATAGTTTGGAGTGAATCACAATCTGTAGTACTTACCCAAAGCAATTTGGAGGAGATATC	6322
Db	3411	ATAGTTTGGAGTGAATCACAATCTGTAGTACTTACCCAAAGCAATTTGGAGGAGATATC	3470
Qy	6323	AGCCTCTGGAAGTACTTCTCTGCAACTGTAAGTCTCCAGATATTCATTAAT	6382
Db	3471	AGCCTCTGGAAGTACTTCTCTGCAACTGTAAGTCTCCAGATATTCATTAAT	3530
Qy	6383	GGCAACCTGTTATATTTGAAATTTTATGAATGCTGAAGTGAAGTGAAGTGAAGTGAAGT	6442
Db	3531	GGCAACCTGTTATATTTGAAATTTTATGAATGCTGAAGTGAAGTGAAGTGAAGTGAAGT	3590
Qy	6443	TCAAACCAATTTATGTTTCTAAATATAGTGTATGTTATGTTGCTGCAATGAAGTGAAGT	6502
Db	3591	TCAAACCAATTTATGTTTCTAAATATAGTGTATGTTATGTTGCTGCAATGAAGTGAAGT	3650
Qy	6503	CAGTCTTAAATACCAATGAGTGAAGGCTGTTCCATGATGGAGAGCTTTGATTTGT	6562
Db	3651	CAGTCTTAAATACCAATGAGTGAAGGCTGTTCCATGATGGAGAGCTTTGATTTGT	3710
Qy	6563	TTTCATAAATCTCTACATTCATCAATCAATTTGAAATTTGAAATTTGATGCTGAGTGAAGC	6622
Db	3711	TTTCATAAATCTCTACATTCATCAATCAATTTGAAATTTGAAATTTGATGCTGAGTGAAGC	3770
Qy	6623	ATTTTGAAGT 6632	
Db	3771	ATTTTGAAGT 3780	
RESULT 7			
AK122241			
LOCUS Mus musculus mRNA for mKIAA0305 protein.			
DEFINITION AK122241			
ACCESSION AK122241			
VERSION AK122241.1 GI:28972138			
KEYWORDS FLI CDNA.			
SOURCE Mus musculus (house mouse)			
ORGANISM Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE 1 Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S., Nakajima, D., Nagase, T., Ohara, O. and Koga, H.			
TITLE Prediction of the coding sequences of mouse homologues of KIAA gene. II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries			
JOURNAL DNA Res. 10, 35-48 (2003)			
REFERENCE 2 (bases 1 to 5438)			
AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.			
TITLE Direct Submission			
JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan			
(E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)			

COMMENT		The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing; Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.	
FEATURES		Location/Qualifiers	
source		1..5438	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/db_xref="taxon:10090"	
		/clone="mbg08215"	
		/tissue_type="brain"	
		/dev_stage="adult"	
		/note="vector:modified pBC SK+"	
gene		1..5438	
		/gene="mKIAA0305"	
CDS		<109..4719	
		/note="CDS is predicted by in silico analysis. Start codon is not identified."	
		/evidence="not experimental"	
		/product="mKIAA0305 protein"	
		/protein_id="BAC65523.1"	
		/db_xref="GI:28972139"	
		/translation="VLSKPTGRMDSYFKAAYSGLDKLLDDPEONPKDYLPDAVAFN QCVSSESASQLALLSKDQRCISTCASSEACEDANETLEGIKEGLTSRPNKKNV AGLDLSVSADASTDEIQSCWRRCKFVCDLISDMGLNVLVHATNSEEDIKQLPDPK ASADLTITLDSVSEALTSSVDCGNAVREEQNNINAGIKNRDNISKEUGKVDMA LPDCKYRNENLKKIISNELETVDFFDMPSVLMKESEMENKTPKELPCCLLK DQCLAEKVAVANNTCELBEGGSNTIAMPKLPENEGISPSDPSAKDENKFLPFD PLOENRSTVEMKQTKVEDSRNLDKNNDIIVHSGDDVPPSLCLSLGSKGSLIHN NEHSLLPNESEGGONDAVLIHESIOKSDVLDGTDLSKCTCSIPLQPVNKKKG QKVEVEVMISGESLESDEDAASAAASPVASVPEAPGCEGTTPSSDMWGQ ELDFYFNIDESMRSGILISDAELDAFLKEOCLSNSTMSAGENVNDSQLMQOITWKL EUDYFNIDESMRSGILISDAELDAFLKEOCLSNSTMSAGENVNDSQLMQOITWKL HDENAGDIYFAEAGAGENGVCETSDKENENGLSIGEXGAIPIERELASCPQ DIRDLVPISIKTQAVGARPKQLSLPFTSRSSKELNKPDVVDVPESEPTANATV STCSADHTPDQSVSPNSNVDIENFDEGSSFTANKDSLKENKESLVLQKQFTW VPSAPCMNCQVKTFTTTRHRHACGKFCVGCNCRKCKLOYLEKEARVCVICYE TINKQAQERWMSPGSCLKNHNECATDPOLOSTQTSPTSPPLTISALKQNVVE GPCSEKQKRFADGILPNEGVADITKLSSSKRSKSDPSPVLPVPMVPMKVDRTS PTEPKNGRLGDIIRSEISQSPCTHAPVERLPGNTGTEGLEPMPGPTLDDVFDSE BESTPTVPANSLPVASISDYRLICSAVKCVNNISLLPDDDIGLPPLATSGEDGS VPVQVERPSHEQIILLBEGEPFPATFVLNANLVNKLVLVSYSSKWSYFSLNGHLG GOAEIIVLLQCLPNEDTVPKDFRIFTITYDKALGKYIENLDNLTFTESFLNKDHG GFLFITPTFONLDGLPVRSPFLGILIQKLEIIPWAKVPMRLMLRGAAYKAYAPL TSVRGRKPLFGEIGHTINLVLDNRYOYTLHNDIQLIHMENKSCIKIPKYSVDV MKVTHSSNEHVISIGASSTEADSHLVQSGDGVYQTCANSATGPKRVTGASVVDV DICKVDSSEERKNKGVISVDGMSVEGFPSEKIGETDETEBEKTVKCTEVFPLKDDQ ISILSSSYQFAKEIAVACSAALCPHLRTLKSNRMNKIGLRVSDIDTMVFEQACGSQL LPQHYLNDLDSALIPVINGGTSNSSLPLEIELAFLENLSE"	
ORIGIN		Query Match 50.3%; Score 3337.4; DB 10; Length 5438; Best Local Similarity 78.9%; Pred. No. 0; Matches 4303; Conservative 0; Mismatches 1021; Indels 133; Gaps 22;	
Qy	208	AGGCATACAGAAATTAATTTCTGAATAAGTCTGCAGGTAGGATGGACAGTATTTTAAAG	267
Db	92	AGCATAGAGAAATGAGTTCAGTAACCCACAGGAGGATGGACAGTACTTTTAAAG	151
Qy	268	CAGCTGTCACTGAGTCTGGACAAACTCTTCTGATGATTTTCAACAGACCAGATGAACAAG	327
Db	152	CAGCTGTCACTGAGTCTGGACAAACTCTTCTGATGATTTTGAAGCAGATCCAGATAAGCAAG	211
Qy	328	ATTATCTCGAGATGTACAAAATGCATATGATTTCAACCACTGCTCAGTCTTCTCAGAGT	387
Db	212	ATTATCTCCAG-----ATGCATATGCTTTTAAACCACTGCTCAGTCTTCTCAGAGT	262
Qy	388	TGGCTTCTCTCACAGCAACTTCAATGCTCCCAAAGCAACCAAGAGTCGTTAATAGTTGTG	447
Db	263	CGGCTTCTCACAGCTAGCT---TTACTGTCAAAGGACCAAGATGCATCAGTACTTGTG	319

RTNLDKDIISLETVDFMPVSLVMEQSEMSNTKDNPOYKRLPCELLKDDGLAE
 KVAVANNTECLEEGSSNTIAMPKLPENEGSPSDPASKOENFXPLPDPLOERNTS
 VFNKQTVKEDNRNLDKNDNDIIVHSGDDVPVPSLSGLSLGSLGSLHNEHSDILEP
 PNESEGONDADVTHEEIOKSDVLDGETDLSKKTCSIFLOPVNEKKEGVEVEEM
 VIGSLESPEDASALAAAGSPVALSAASVPEAPGPCGLTPSPSDMDGOELDENID
 ESWRSGLISDAELDAFLKEOCLSNSNTMSAGENVDSOLQXNQIITMKGLHNDAGDI
 YFNAAGAAGENGVCNCHTSKXENTENNGLSIGERGAIPTERELSAACOPDIRDELVP
 PSIKTAGVGARPKQLSLUPPGPRSRKLNKLPDVPDPSEPCATANATVSCSADHI
 PDSQVFNYSYDIENFEDGSFVANKDLSLENKRKESLVIGQKOPTWPDSEAPN
 CMNCQVFTFKRHHRCACGKVFQVCCNRKQLYLEKEARVCVICYETINKQAQF
 ERNMSGKSLKHNHNECATDQLOBTOTSTPSPTTLPISALQKQVNEVGPCKEQK
 RVFADGLPENGVAOTTKLSSGKSCDDSPVLDPVMTINKVDRHSPTEKPN
 GLGDIIRSELSPVCHTAPVESLPGNTGTEGLPMGPBETLSDVDVDEEESTPTV
 PANSGLPVASISDYRLCSVAKVCVNNISLLPDDDIGLPLLATSEDSQVWQERP
 SHEQIILLSEGGEPFATFVLNANLVNKLKYLISSEKYWFSTNGHUGQBLIITP
 LOCLPNEDVPKDIIFLFTIYKDALKGKYLENDLNTFTESLNSKHGGFLIITP
 FQNLGDLVPSPFLCILLQKLEIPWAKVFPNMLRLGAELYKAPALPSVGRGRK
 LFEIGHTIMNLLVDNRNOYTLHNDIQLLHMENKSCIKIPKRYKSDVMKVIHSSN
 EHVISGASPEADSHLVQSDGVYTOANSATGPKRVKTVGASVFNENGALKTSGG
 FLAKSSIVSDGLVQITPEYMEGLRLALREKQDFRLOCCKVGDADVLREXVDLCWVDS
 ERNKGVISSVDGNSVEGFESEKIKLETDTEEEKIVKCTEVEYFYLKDDDISILSSY
 QFAKEIAVACSALCPHLRLTKSNRMKIGLRVSDTDMVFOAGCEGQLLPOHYLND
 LDSALIPVHGGTSNSLPLEIEBLAFPILENLSE"

misc_feature

2337..2531

/note="FYVE; Region: FYVE zinc finger. The FYVE zinc
 finger is named after four proteins that it has been found
 in: Fab1, YORF/ZK632.12, Vac1, and BEAL. The FYVE finger
 has been shown to bind two Zn++ ions. The FYVE finger has
 eight potential zinc coordinating cysteine positions. Many
 members of this family also include two histidines in a
 motif R+HHC+XCG, where + represents a charged residue and
 X any residue. We have included members which do not
 conserve these histidine residues but are clearly related"
 /db_xref="CDD:pfam01363"

ORIGIN

Query Match 50.3%; Score 3333; DB 10; Length 5480;
 Best Local Similarity 78.8%; Pred. No. 0;
 Matches 4298; Conservative 0; Mismatches 1020; Indels 133; Gaps 22;
 QY 208 AGGCATACAGAAATAAATCTGTAATAGTCTGGAGTAGATGGACAGTATTTTAAAG 267
 DB 109 AGACATAGAAATAGTGTCTAGTAAGCCACAGAGGAGGTGACAGTACTTTTAAAG 168
 QY 268 CAGCTGTGAGTACTGACAACTCTTGATGATTTTGAACAGACCAGATGAACAAG 327
 DB 169 CAGCTGTGAGTGTGACAACTCTTGATGATTTTGAACAGATCCAGATAGCAAG 228
 QY 328 ATTAATCTCGAGATGTACAAAATGCATATGATTTCAACCTGCTCAGTTTCTTCAGAGT 387
 DB 229 ATTAATCTCCAG-----ATGCATATGCTTTTAAACCAGTGTCTCAGTTTCTTCAGAGT 279
 QY 388 TGGCTTCTCAGCGGACTTCATGTCTCCAAAGACCAAGAGTGCCTTAATAGTTGTG 447
 DB 280 CGGCTCTTCCACAGCTAGCT---TTACTGTCAAAGGACCAAGATGATCAGTACTTTGTG 336
 QY 448 CCTCATCAGAAACAAGGTATGGA---ACAAATGAGAGTTCCTCGAATGAAACCACTCA 504
 DB 337 CCTCATCAGAGCATGCTGTGAAGATGCAATGAGACTTTTCTTGAAGGAAGATACACG 396
 QY 505 AGGAGACTTACTTATACAAAATGAAAAAATGTAAACAGACTGTGATCTTCTTCTCTG 564
 DB 397 AGGCTTTAACCTCTAGACCGAATGAAAAAACGTTAGCGGAGTGTGATCTTCTTCTCTG 456
 QY 565 TGGATGGTGGTACTTTCAGATGAAATCCAGCCGTTTATATATGGGACGATGTAGTAAACCTA 624
 DB 457 TGGATGGAGTACTTTCAGATGAGATCCAGCTTCTCTGTATGAGACGCTGTAGTAAAGCTG 516
 QY 625 TCTGTGATCTGATAGTACATGGGTAACTTAGTTCATGCAACCAATAGTGAAGAGATA 684
 DB 517 TCTGTGACCTGATAAGTGACATGGGTAACTTAGTTCATGCAACCAATAGTGAAGAGACA 576
 QY 685 TTNAATAATTATTCAGAGATGATTTTAAAGTCTAATGAGATTCCTGATTTGGATTGGATT 744

DB 577 TTAAAGCAGTTATGTCAGATGACCCCTAAGTCTTCTGCAAGTACTTTGATCACTTGGACT 636
 QY 745 TATCTTTCAGTGTGAGATGACTCCCTGTGTGTCTTCAACAGACCATGATAGTATGTCAT 804
 DB 637 CGTCTTCAGTTTCAGAGGCTCTCACTGTCTCTTCACTAGTACTGTGTTAGTAACTGTCA 696
 QY 805 GAGAACACAGNATGATATCAGTCTGTAATTTACAAAATAGAGAAATCGAGGAATCAAG 864
 DB 697 GAGAAAGACAGAAATACATCAACCGTGTGTATAAAAAAGAGAGATAT---TAGTATCAAG 753
 QY 865 AATTCGGGTATAAAACTAGATACAAACATTTTCAAGTCTCTATAAATTACAGTGGAAACAGAAA 924
 DB 754 AATTCGGGTATAAAACTAGATGACATGCGACATTTTGAATCTTGTGTAATATAATAGAACAGAAA 813
 QY 925 AATTTAAAGATAAAAAGATCTTTAATCAGTGTAGATCAATTTGTGATTTTAACTATGTCAT 984
 DB 814 AATTTAAAGATAAAAATTTAATTAATAGTGTAGAAAACA---GTTGATTTTTCAGATGCCAT 870
 QY 985 CTGCTTTGACTCGACAAAGTTTCCAAAATGTTTTCATGCGAAAGACAAAGTGTACAAACAAGA 1044
 DB 871 CTGTTTGTGATGGAACAAAGTTTCTGAAATGTCTAATACCAAGATTAACCCCAATACAAGA 930
 QY 1045 GCCAGCCATGTGGATTACTAAAAAGATTTGGCTTGTAGTAAAAAGAGAGTGTAGTGGCAG 1104
 DB 931 GGCTGCCATGTCAATTGTTTAAAGATGATGGCTGTTTGGCAGAGAGAAAGTAGCTGTGG 990
 QY 1105 TCATAACTGCCCGACAGATGTTTAAAGAGAGAGG-----GCAAGACAGTGTCTTGAACCT 1158
 DB 991 CAGTCAATAAATACAGAGTGTTTTAAAGAGAGAGGAGGAGGATACCAATTTGCTATGCCCT 1050
 QY 1159 GCAGCTTCCGAAAAAATGAAGATTTTATGCTTAAATGATTTCAAAATTCAGAGATGAAAAAT 1218
 DB 1051 GCAGCTTCCGAAAAATGAAGGATATCCCAAGTGTGCCAGCTTCAAAAGATGAAAAAT 1110
 QY 1219 TCATAATCTGACTTTTCCCTTTCAGGAGATGAAGCT-----GTTATAAACATCTG 1272
 DB 1111 TCAAGTTACTGACTTTTCCCTTCCAGAGAAATAGAACTTCTGTGTCTATGAACAGAGCTG 1170
 QY 1273 CACAAGAGACTCAAAAAGTTTGTAGACCTTAAAGGATAATGATGATAATCCAAGATTCCTCTT 1332
 DB 1171 TAAAGAGACTCAAGAAATTTAGACCTTAAAGATAATTAATGATAT----- 1216
 QY 1333 CAGCTTTACATGTTTCCAGTAAAGATGTCGGTCTCTCAATGTCCTGTCTTCTCGGTCTG 1392
 DB 1217 ----AGTCCATGTTTCAGGTGATGATGTGCCCTTCACTGTCTGTCTTTCATTTGCTG 1272
 QY 1393 GGTCTATGTGTGATCATTAATTTGAAGTAAAGACGGGGTGTATTTTACCTCAGCATG 1452
 DB 1273 GGTCTTTGTGTGATCGTTAATTTCAATAATTAATGAACACAGTGTATTTTACCTCCTAATG 1332
 QY 1453 AACATAAAGATTAATATACAAGATGCAAGTGTGACTTATACATGAAGAAATACAGAAACAGTGTG 1512
 DB 1333 AGTCTGAAGGCCAGAACAAATGATGAGTGTGACTTATACATGAAGAGATACAGAAAGTGTGATG 1392
 QY 1513 TTCTAGTGGGAAACCAATTCAGAAAGATGATCTTTTGAACAGGAGAAATGTAAAGCA 1572
 DB 1393 TTTTAGATGGG-----GAGACAGATCTTTTCAAAAGAGGAAACGCTGTAGAGCA 1440
 QY 1573 TACTCCTTCAGTCAATTAATTTGAAGGATGGAAGACAGAGAAAGATAGATCCTGACCAAGACAG 1632
 DB 1441 TATTTCTTCAGCCAGTAAATGAAGAGAGGGAGGAAAGGTTGAGGTTGAGGAGATGG 1500
 QY 1633 TAATCAGAGCTGAGTCTTTTGA-----TGGTGGTGAACCAAGTCTTACAGTTGTAGAAAT 1686
 DB 1501 TAATCAGTGGTGTGAGTCTTTGGAGTCCCTGTAGAGAGCCAGCTCTGTGTGAGCAGCTGGGT 1560
 QY 1687 CTCAGAGGGGCTTTCTGCACTCATGTCCAGAGTCTTCTGATTTCTGATTTCTGAGGTTTFA 1746
 DB 1561 CTCGGTGGCACTTTCTGTGCAAGTGTCCCAAGAGCTCTGTGTCCATGTGAAGTCT--- 1618
 QY 1747 TTAATACTTTTCAAGCAATGATATGATGGCAAGACTTAGATTTACTTTAATATGATG 1806


```
/tissue type="human retina"
/clone_lib="686 (synonym: hlcc3) . Vector pSport1_sfi; host
DH10B; sites SfiIA + SfiIB"
/dev stages="adult"
polyA_signal 5117..5122
polyA_site 5140
ORIGIN
Query Match 40.0%; Score 2655.2; DB 9; Length 5158;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2668; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 3939 GACCTTCGAATTAACAGTACCTTGCAATATAGATCAATCACTGTTGATTTCATATGGAA 3998
DB 2473 GACCTTCGAATTAACAGTACCTTGCAATATAGATCAATCACTGTTGATTTCATATGGAA 2532
QY 3999 ATGGGAAAAAGCTGCATAAAAATACACGGAAGAAAGTACAGTGTATGAATGAAAGTACTA 4058
DB 2533 ATGGGAAAAAGCTGCATAAAAATACACGGAAGAAAGTACAGTGTATGAATGAAAGTACTA 2592
QY 4059 AATTCCTCAATGAGCATGTCATTAGCATTCGAGCAAGTTTCAGTACAGAGCAGATTCT 4118
DB 2593 AATTCCTCAATGAGCATGTCATTAGCATTCGAGCAAGTTTCAGTACAGAGCAGATTCT 2652
QY 4119 CATCTAGTCTGTATACAGAAATGATGGAATTTATGAACACAGGCCAACAGTGCACCTGGC 4178
DB 2653 CATCTAGTCTGTATACAGAAATGATGGAATTTATGAACACAGGCCAACAGTGCACCTGGC 2712
QY 4179 CATCTAGAAAGTACAGGTGCAAGTTTCTGTTGTTTCAATGAGCTCTAAAAACATCT 4238
DB 2713 CATCTAGAAAGTACAGGTGCAAGTTTCTGTTGTTTCAATGAGCTCTAAAAACATCT 2772
QY 4239 TCAGGATTTCTTGCTAGCTCAGCATAGTTCAAGATGCTTAATGTTACAAATACCTCCA 4298
DB 2773 TCAGGATTTCTTGCTAGCTCAGCATAGTTCAAGATGCTTAATGTTACAAATACCTCCA 2832
QY 4299 GAGACCATGAATGGCTTCGCGCTAGCTTTACGAGAACAGAAAGTCTTTAAATTTACATGT 4358
DB 2833 GAGACCATGAATGGCTTCGCGCTAGCTTTACGAGAACAGAAAGTCTTTAAATTTACATGT 2892
QY 4359 GGGAAAGTTGATGCTAGCTCAGAGTACCTGAGAAATACCTGGATCTCTGGGTAGTCTGAA 4418
DB 2893 GGGAAAGTTGATGCTAGCTCAGAGTACCTGAGAAATACCTGGATCTCTGGGTAGTCTGAA 2952
QY 4419 GAAAAAGGAAACAAAGGAGTTATCAGTTTCAGTGGATGGAATATCATTTACAGGATTTCCA 4478
DB 2953 GAAAAAGGAAACAAAGGAGTTATCAGTTTCAGTGGATGGAATATCATTTACAGGATTTCCA 3012
QY 4479 AGTGAAGAAATAAACTGGAGCAGATTTTGAACCCGATGAGAGATGTAAATGTACC 4538
DB 3013 AGTGAAGAAATAAACTGGAGCAGATTTTGAACCCGATGAGAGATGTAAATGTACC 3072
QY 4539 GAGGTGTTCTACTTTCTTAAAGGACAGAGATTTATCTATTTATCAACTTCTTATCAGTTT 4598
DB 3073 GAGGTGTTCTACTTTCTTAAAGGACAGAGATTTATCTATTTATCAACTTCTTATCAGTTT 3132
QY 4599 GCAAAAGAAATAGCCATGGCTTTAGTCTGGCTGTGCGCTGCGCTCCACCTGAAACCTCTAAA 4658
DB 3133 GCAAAAGAAATAGCCATGGCTTTAGTCTGGCTGTGCGCTGCGCTCCACCTGAAACCTCTAAA 3192
QY 4659 AGTAATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGACACTGATATGTTGAATTT 4718
DB 3193 AGTAATGGATGAATAAATTTGGACTCAGAGTTTCCATTTGACACTGATATGTTGAATTT 3252
QY 4719 CAGCAGGATCTGAAGGCCAACTTCTGCTCAGCATTTATCTTAAATGATCTTGTATGATGCT 4778
DB 3253 CAGCAGGATCTGAAGGCCAACTTCTGCTCAGCATTTATCTTAAATGATCTTGTATGATGCT 3312
QY 4779 CTGATACCTGTGATCCATGGTGGGACCTCCAGCTAGTTTACCATTTAGAAATAGAAATTA 4838
DB 3313 CTGATACCTGTGATCCATGGTGGGACCTCCAGCTAGTTTACCATTTAGAAATAGAAATTA 3372
QY 4839 GTGTTTTTCATTATAGAACATCTTTTTTTAGTGAAGAAATGTGCCATATTACATATTGCAA 4898
```

```
DB 3373 GGTGTTTTTCATTATAGAACATCTTTTTTATGTAAGAAATGTGCCATATTACATATTGCAA 3432
QY 4899 CTAATATTTGTTAAAACTAAGTCCACACCTAAAGCTGAAATGCCACAAACACTAAAAGTAT 4958
DB 3433 CCTAATTTGTTAAAACTAAGTCCACACCTAAAGCTGAAATGCCACAAACACTAAAAGTAT 3492
QY 4959 AATATGTCGTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGCAAGGAATGATCTTTTCA 5018
DB 3493 AATATGTCGTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGCAAGGAATGATCTTTTCA 3552
QY 5019 AATCATTTAGCACAAATATTAAATATCTAAAAATTTTAAAGAGATCCATCTTTCTGTAGCTT 5078
DB 3553 AATCATTTAGCACAAATATTAAATATCTAAAAATTTTAAAGAGATCCATCTTTCTGTAGCTT 3612
QY 5079 TACAATTAATTTAAGTACTTAAAGACACAGGATTTCTTTTAAAGAAATTTATAGCATTTAC 5138
DB 3613 TACAATTAATTTAAGTACTTAAAGACACAGGATTTCTTTTAAAGAAATTTATAGCATTTAC 3672
QY 5139 TGTGTTATTTAAATGCTAAGCCAAAGTATCTGCACTTAGGTATACCTCTTTATGCCAATA 5198
DB 3673 TGTGTTATTTAAATGCTAAGCCAAAGTATCTGCACTTAGGTATACCTCTTTATGCCAATA 3732
QY 5199 ATGATTTTAAATGAAGCTCTTTTCAAGATGTAACTTTTGAAGAAATATCTGCTTTGTGT 5258
DB 3733 ATGATTTTAAATGAAGCTCTTTTCAAGATGTAACTTTTGAAGAAATATCTGCTTTGTGT 3792
QY 5259 ATATGCCAGTTAGATACCTGTTTCTTAAAGTCTGCAAAATGTTTTCAGTGGGACAAAA 5318
DB 3793 ATATGCCAGTTAGATACCTGTTTCTTAAAGTCTGCAAAATGTTTTCAGTGGGACAAAA 3852
QY 5319 ACCAGTTTTGAGGCTCTTAGACTTATAATCTTTTGAATAAACTGATACTTTATTTGTATA 5378
DB 3853 ACCAGTTTTGAGGCTCTTAGACTTATAATCTTTTGAATAAACTGATACTTTATTTGTATA 3912
QY 5379 ATGAGTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTTATATATCAAGATT 5438
DB 3913 ATGAGTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTTATATATCAAGATT 3972
QY 5439 TTGCTCTTTTCTCTCATAATTTATACATATGATGTATATATATATCCACATATAGTT 5498
DB 3973 TTGCTCTTTTCTCTCATAATTTATACATATGATGTATATATATATCCACATATAGTT 4032
QY 5499 TTCCCTGTATTAATGGATTTAAATAATTTGCGGTGCTTCAGGACTTTTTCCTCTATA 5558
DB 4033 TTCCCTGTATTAATGGATTTAAATAATTTGCGGTGCTTCAGGACTTTTTCCTCTATA 4092
QY 5559 TTTAAGTATATTTGTTTATAGCAAGAACATATTTCTGAATGTTTATATAAATCTTTAATAA 5618
DB 4093 TTTAAGTATATTTGTTTATAGCAAGAACATATTTCTGAATGTTTATATAAATCTTTAATAA 4152
QY 5619 TTTATATGTAAGTAAATTTTTTGTATCAATGCAATATTTTTTTTCTCCTCTCTCTCC 5678
DB 4153 TTTATATGTAAGTAAATTTTTTGTATCAATGCAATATTTTTTTTCTCCTCTCTCTCTCC 4211
QY 5679 AAACATATACCTGTATTTTACCACTTCTAAGAGTCACTGACGACGGGCGCAGATGACCTT 5738
DB 4212 AAACATATACCTGTATTTTACCACTTCTAAGAGTCACTGACGACGGGCGCAGATGACCTT 4271
QY 5739 GAAGTAGTCATTATGTAGCAATTAATGAAGCCTGAAAACAGGTTTTTTTACCTCTCTTA 5798
DB 4272 GAAGTAGTCATTATGTAGCAATTAATGAAGCCTGAAAACAGGTTTTTTTACCTCTCTTA 4331
QY 5799 ATCCCTTAGAAATTTCTTGGCACTTGGCATATTTTTCATTGACCTGGTGTATAGTATAA 5858
DB 4332 ATCCCTTAGAAATTTCTTGGCACTTGGCATATTTTTCATTGACCTGGTGTATAGTATAA 4391
QY 5859 ATTAAATGAACTAAATTTACTTTTGTGATATTTTAAATTTCTTTATATGTTAGTTATTTTA 5918
DB 4392 ATTAAATGAACTAAATTTACTTTTGTGATATTTTAAATTTCTTTATATGTTAGTTATTTTA 4451
QY 5919 TAAACAGGATTTAAACATAAGTTTAAATCTTATGATTTTGAATTTTGTATAGAGCTTTCTC 5978
```

Db 4452 TAACAGGATATTAACATAAGTTAAATCCTATGTATTGAAATTTGTAATGTTTACAGAGCTTTCCTC 4511
 Qy 5979 TTTTACTTCAACACAGCAAAAAGTGGGGGCGATATTTAGTCCCTGTCATTTTAAGTTATGTA 6038
 Db 4512 TTTTACTTCAACACAGCAAAAAGTGGGGGCGATATTTAGTCCCTGTCATTTTAAGTTATGTA 4571
 Qy 6039 AAAAATTTAATCATTTATTTGATGCTTTTAAACATTTCTCATGTGTAATATATGTTTTGTA 6098
 Db 4572 AAAAATTTAATCATTTATTTGATGCTTTTAAACATTTCTCATGTGTAATATATGTTTTGTA 4631
 Qy 6099 TCAAAAACATCATATATTTTCAAGAAAAGAAATATTTGTAATAGCCCTGTTTTAAGAA 6158
 Db 4632 TCAAAAACATCATATATTTTCAAGAAAAGAAATATTTGTAATAGCCCTGTTTTAAGAA 4691
 Qy 6159 AAATATTTTGAAGCATCTCAACTTTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 6218
 Db 4692 AAATATTTTGAAGCATCTCAACTTTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 4751
 Qy 6219 GTCTCAAGCTAGGAGACATGAGAAATTTTAAATCAGTTTGGGCATATAGTTTGGACTGAAT 6278
 Db 4752 GTCTCAAGCTAGGAGACATGAGAAATTTTAAATCAGTTTGGGCATATAGTTTGGACTGAAT 4811
 Qy 6279 CACATCTGTAGTACTTACCAAGACAAATTTGGAGGAGATATCAGCCTTCTGGAAGTAG 6338
 Db 4812 CACATCTGTAGTACTTACCAAGACAAATTTGGAGGAGATATCAGCCTTCTGGAAGTAG 4871
 Qy 6339 CTACTTCTGCAACAATGTAAGTGTGCGAGATATTCATTAATGGAATGGAACCTGTTTAAAT 6398
 Db 4872 CTACTTCTGCAACAATGTAAGTGTGCGAGATATTCATTAATGGAATGGAACCTGTTTAAAT 4931
 Qy 6399 TTGTGAATTTATTTGAATGGTGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 6458
 Db 4932 TTGTGAATTTATTTGAATGGTGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 4991
 Qy 6459 TTTTCTAAATATAGTATGATGATCTGCAATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6518
 Db 4992 TTTTCTAAATATAGTATGATGATCTGCAATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5051
 Qy 6519 AAATGTTAGAGGGCTGTTCCATGATGGAGAGCTTTGGATTTGTTTTCATATAAATCTCTA 6578
 Db 5052 AAATGTTAGAGGGCTGTTCCATGATGGAGAGCTTTGGATTTGTTTTCATATAAATCTCTA 5111
 Qy 6579 CATTCAATAAATAATGGAATTAATGTCCTGAA 6610
 Db 5112 CATTCAATAAATAATGGAATTAATGTCCTAAA 5143

RESULT 10
 BC030808
 LOCUS
 DEFINITION Homo sapiens zinc finger, FYVE domain containing 16, mRNA (CDNA clone MGC:22300 IMAGE:4701661), complete cds.
 ACCESSION BC030808
 VERSION BC030808.1 GI:21411214
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2902)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S., Carninci,P., Frange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villaen,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,

Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 2902)
 Strausberg,R.
 Direct Submission
 Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 37 Row: a Column: 16
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7662047.

FEATURES

1..2902
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:22300 IMAGE:4701661"
 /tissue_type="Lung"
 /clone_lib="NIH MGC_77"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 1..2902
 /gene="ZFVVE16"
 /note="synonyms: KIAA0305, ENDOFIN"
 /db_xref="LocusID:9765"
 111..2540
 /codon_start=1
 /product="ZFVVE16 protein"
 /protein_id="AAH30808.1"
 /db_xref="GI:21411215"
 /db_xref="LocusID:9765"
 /translation="MDSYFKAASVSLDLDLDFEQNPQDYLDQVQVAYDSNHCVS
 SLSASORTSLPKDQECVNCASSETSYGTNSSLNKKTLKLTSLGTSIQENKVTGLDL
 LSSVDGTTSDTEIQLYMGRCSPICDLISDMGNLVHATNSREDIKLLPDPFKSNADS
 KTVIKQAQEDSKSLDKNDVIQSSALHVSXKVPSSISCLPASMGCSLIESK
 ARGDLPOHEKNDIQDAVTHEIQNSVWLGEPFKENDLLKQCKSILLQSLIEG
 MEDRKIDPQTVIRAESLDGDTSTVVESSQGLSGTHVPSSDCEGFIINTFSSNDM
 DQGLDYNIDEGASGLISDAELDAETEQYLQTTNIKSPRENVDSKSMQOIDM
 KGLDGNINIFYNFAAGAGSHGINICIVDKQNTIENGLSGEKSTIPVQGGLP
 TSKRITNQLSVDSVSGAPKQLFSLPSRTRSKNDLKNKPDVPTIEGSPAD
 TVPTITCDASTDAPQVFSNGNYDIESNQSFSVTANEDSVPTCKEGLVGQK
 QPTWYDSEAPNMCNQKFTPTKRRHRCACGKVFVCGCCNRKCKLQYLEARVCV
 VCYETISKGEK"
 2334..2528
 /note="FYVE; Region: FYVE zinc finger. The FYVE zinc
 finger is named after four proteins that it has been found

Gene

CDS

misc_feature


```
QY 2165 TACAGTCAATCTGTTGGAGGGCCAGACCTAAGCAATTTAGCTTCCATCAAGAAC 2224
Db 2027 TAACAGTCAATCTGTTGGAGGGCCAGACCTAAGCAATTTAGCTTCCATCAAGAAC 2086
QY 2225 AAGAGTTCAGACCTGAATAGACAGATGTTCCAGATACATAGAGAGTGAACCCAG 2284
Db 2087 AAGAGTTCAGACCTGAATAGACAGATGTTCCAGATACATAGAGAGTGAACCCAG 2146
QY 2285 CACAGCAGATACCGTTGTTCCCAATCACTTGTGTATAGATTCTACAGCTGATCCACAGGT 2344
Db 2147 CACAGCAGATACCGTTGTTCCCAATCACTTGTGTATAGATTCTACAGCTGATCCACAGGT 2206
QY 2345 TAGCTCAACTCAATTAATACATAGATAGAAAGTAATCTGAAGGTGGATCTAGTTTGGT 2404
Db 2207 TAGCTCAACTCAATTAATACATAGATAGAAAGTAATCTGAAGGTGGATCTAGTTTGGT 2266
QY 2405 AACTGCAAAATGAGATTCTGTAACCTGAACACCTTGCACAAAGAGGCTTGGTTTGGGCCA 2464
Db 2267 AACTGCAAAATGAGATTCTGTAACCTGAACACCTTGCACAAAGAGGCTTGGTTTGGGCCA 2326
QY 2465 GAAACAGCTTACTGGGTTCTCTGATTGAGAGTCCAACTGATGAACTGCCAAGTCAA 2524
Db 2327 GAAACAGCTTACTGGGTTCTCTGATTGAGAGTCCAACTGATGAACTGCCAAGTCAA 2386
QY 2525 ATTACTTTTACCAAGCGGCACACCATTCGCGAGCATGTCGGAAGTATTTTGGTGT 2584
Db 2387 ATTACTTTTACCAAGCGGCACACCATTCGCGAGCATGTCGGAAGTATTTTGGTGT 2446
QY 2585 CTGTTGTAATAGGAAGTGAACCTGCAATATCTAGAAAGAGCAAGAGTATCTGTAGT 2644
Db 2447 CTGTTGTAATAGGAAGTGAACCTGCAATATCTAGAAAGAGCAAGAGTATCTGTAGT 2506
QY 2645 CTGCTATGAACTATTAATAGTAAAGTCTAG 2672
Db 2507 CTGCTATGAACTATTAATAGTAAAGTCTAG 2534
```

RESULT 11

```
AC008382/c
LOCUS 252015 bp DNA linear PRI 23-OCT-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-218P2, complete sequence.
ACCESSION AC008382
VERSION AC008382.6 GI:16328245
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 252015)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 252015)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 252015)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 23, 2001 this sequence version replaced gi:9255963.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.7.
Location/Qualifiers
1 . 252015
/organism="Homo sapiens"
/mol_type="genomic DNA"
```

```
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-218P2"
```

ORIGIN

```
Query Match 34.0%; Score 2254.4; DB 9; Length 252015;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2258; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 316 CAGATGAACAAGATTATCTCGCAGATGTACAAAATGTCATATGATTTCTAACCACTCTCTCAG 375
Db 20953 CAGATGAACAAGATTATCTCGCAGATGTACAAAATGTCATATGATTTCTAACCACTCTCTCAG 20894
QY 376 TTCTTCAGAGTGGCTTCTCCACAGCGAATCTTCTATCAAAAATGAAAAAATGTAACAGAGTTCCCTGAAATGAA 495
Db 20893 TTCTTCAGAGTGGCTTCTCCACAGCGAATCTTCTATCAAAAATGAAAAAATGTAACAGAGTTCCCTGAAATGAA 20834
QY 436 TTAATAGTTGTGCTTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTGAAATGAA 20774
Db 20833 TTAATAGTTGTGCTTCATCAGAAACAAGCTATGGAACAAATGAGAGTTCCCTGAAATGAA 20774
QY 496 AAACACTCAAGGAGCTTACTTCTATCAAAAATGAAAAAATGTAACAGAGTTCTGATTC 555
Db 20773 AAACACTCAAGGAGCTTACTTCTATCAAAAATGAAAAAATGTAACAGAGTTCTGATTC 20714
QY 556 TTCTTCAGAGTGGCTTCTCCACAGCGAATCTTCTATCAAAAATGAAAAAATGTAACAGAGTTCCCTGAAATGAA 615
Db 20713 TTCTTCAGAGTGGCTTCTCCACAGCGAATCTTCTATCAAAAATGAAAAAATGTAACAGAGTTCCCTGAAATGAA 20654
QY 616 GTAAACCTATCTGTGATCTGATAAGTGCATGGGTAACTTAGTTTCATGCAACCAATAGTG 675
Db 20653 GTAAACCTATCTGTGATCTGATAAGTGCATGGGTAACTTAGTTTCATGCAACCAATAGTG 20594
QY 676 AAGAAGATATTAATAAATTTATGCGCAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATG 735
Db 20593 AAGAAGATATTAATAAATTTATGCGCAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATG 20534
QY 736 GATTGATTATCTTCAGTGTGATGATCTTCTCAACAGACCAATGATAGTG 795
Db 20533 GATTGATTATCTTCAGTGTGATGATCTTCTCAACAGACCAATGATAGTG 20474
QY 796 ATACTGTGAGAAACAACAGAAATGATATCAGTTCTGAAATACAAAATGAGAAATCGGAG 855
Db 20473 ATACTGTGAGAAACAACAGAAATGATATCAGTTCTGAAATACAAAATGAGAAATCGGAG 20414
QY 856 GAATCAAGAAATTTAAAGATGATAGTAAAGTGTCTTCAACAGACCAATGATAGTG 915
Db 20413 GAATCAAGAAATTTAAAGATGATAGTAAAGTGTCTTCAACAGACCAATGATAGTG 20354
QY 916 GAACAGAAATTTAAAGATGATAGTAAAGTGTCTTCAACAGACCAATGATAGTG 975
Db 20353 GAACAGAAATTTAAAGATGATAGTAAAGTGTCTTCAACAGACCAATGATAGTG 20294
QY 976 ACATGTCATCTGCTTTGACTCGACAAAGTTCGAAATGTTTCATGCCAAGCAAGCTAC 1035
Db 20293 ACATGTCATCTGCTTTGACTCGACAAAGTTCGAAATGTTTCATGCCAAGCAAGCTAC 20234
QY 1036 AACACAGAGGCGCAGCATGTGGATTACTAAAAGATGTTGGCTTAGTAAAGAGGAGTAG 1095
Db 20233 AACACAGAGGCGCAGCATGTGGATTACTAAAAGATGTTGGCTTAGTAAAGAGGAGTAG 20174
QY 1096 ATGTGCGAGTCAATACTGCGCAGAAATGTTTAAAGAGAGGCGGCAAGCAAGTCTTTGA 1155
Db 20173 ATGTGCGAGTCAATACTGCGCAGAAATGTTTAAAGAGAGGCGGCAAGCAAGTCTTTGA 20114
QY 1156 CTGCGAGCTTCCGAAATGAAATTTATGCTTAAATGATTCAAATCAAGAGATGAA 1215
Db 20113 CTGCGAGCTTCCGAAATGAAATTTATGCTTAAATGATTCAAATCAAGAGATGAA 20054
QY 1216 ATTTCAAAATTAACCTGACTTTTCTTTTCCAGGAAAGTAAAGCTGTTTAAACCAATCTGCAC 1275
Db 20053 ATTTCAAAATTAACCTGACTTTTCTTTTCCAGGAAAGTAAAGCTGTTTAAACCAATCTGCAC 19994
```

QY 1276 AAGAGACTCAAAAGTTTACAGCTTAAGGATAATGATGTAATCAAGATTCCTCTTCAG 1335
 Db 1993 AAGAGACTCAAAAGTTTACAGCTTAAGGATAATGATGTAATCAAGATTCCTCTTCAG 1934
 QY 1336 CTTTACATGTTTCCAGTAAAGATGCGGCTCCTCATGTCCTGCTTCCTCGGCTCGGT 1395
 Db 1933 CTTTACATGTTTCCAGTAAAGATGCGGCTCCTCATGTCCTGCTTCCTCGGCTCGGT 19874
 QY 1396 CTATGTTGGATCATTAATTCAGAGTAAGACACGGGCTGATTTTACCTCAGCATGAAC 1455
 Db 19873 CTATGTTGGATCATTAATTCAGAGTAAGACACGGGCTGATTTTACCTCAGCATGAAC 19814
 QY 1456 ATAAAGATAATATACAGATGCAGTACTATACATGAAGAAATACAGAACAGTGTGTTTC 1515
 Db 19813 ATAAAGATAATATACAGATGCAGTACTATACATGAAGAAATACAGAACAGTGTGTTTC 19754
 QY 1516 TAGTGGGGAAACCATTCAGAGATGATCTTTTGAACACAGGAATAATGTAAGCATAC 1575
 Db 19753 TAGTGGGGAAACCATTCAGAGATGATCTTTTGAACACAGGAATAATGTAAGCATAC 19694
 QY 1576 TCCTTCAGTCAATTAATTCAGAGGATGGAAGACAGAAAGATAGATCTCTGACACAGACATA 1635
 Db 19693 TCCTTCAGTCAATTAATTCAGAGGATGGAAGACAGAAAGATAGATCTCTGACACAGACATA 19634
 QY 1636 TCAGAGTCAATGTTGGATGGTGGACACACAGTCTCTACAGTTGTAGATCTCAAGAGG 1695
 Db 19633 TCAGAGTCAATGTTGGATGGTGGACACACAGTCTCTACAGTTGTAGATCTCAAGAGG 19574
 QY 1696 GCGTTCTGGCACTCATGTCACAGCTCTCTGATGTTGTGAAGGTTTATTAACTT 1755
 Db 19573 GCGTTCTGGCACTCATGTCACAGCTCTCTGATGTTGTGAAGGTTTATTAACTT 19514
 QY 1756 TTTCAAGCAATCAATGATGGCAAGACTTAGATCTTTAATTAATGATCAAGGCGCAA 1815
 Db 19513 TTTCAAGCAATCAATGATGGCAAGACTTAGATCTTTAATTAATGATCAAGGCGCAA 19454
 QY 1816 AAGTGGCCCACTAATGATGATGCTGACACTGATGCTTCTGACAGACAGTACTTC 1875
 Db 19453 AAGTGGCCCACTAATGATGATGCTGACACTGATGCTTCTGACAGACAGTACTTC 19394
 QY 1876 AGACACTAAACATAAGTCTTTTGAAGAAATGTAATGATCTTAAATCGCAATGAATC 1935
 Db 19393 AGACACTAAACATAAGTCTTTTGAAGAAATGTAATGATCTTAAATCGCAATGAATC 19334
 QY 1936 AGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATATTTCAATGCAGAG 1995
 Db 19333 AGATAGATATGAAGGCTTAGATGATGGAACATCAATATATATATTTCAATGCAGAG 19274
 QY 1996 CAGGAGCTATTGGGAAAGTCACTGTTAATATAATTTGGAACACAGTTGATAACAAA 2055
 Db 19273 CAGGAGCTATTGGGAAAGTCACTGTTAATATAATTTGGAACACAGTTGATAACAAA 19214
 QY 2056 ATACAATAAGAAATGGGCTTTCTTTAGAGAAAAAGACATATTCAGTTCAACAGGT 2115
 Db 19213 ATACAATAAGAAATGGGCTTTCTTTAGAGAAAAAGACATATTCAGTTCAACAGGT 19154
 QY 2116 TACCTACAGTAAGTCTGAGATTAACAATCAATATCACTCTCTGATTAATTAACAGTCAAT 2175
 Db 19153 TACCTACAGTAAGTCTGAGATTAACAATCAATATCACTCTCTGATTAATTAACAGTCAAT 19094
 QY 2176 CTGTTGGAGGGCCAGACCTTAAGCAATTTAGCCCTTCATCAAGAAACAGGATTCAA 2235
 Db 19093 CTGTTGGAGGGCCAGACCTTAAGCAATTTAGCCCTTCATCAAGAAACAGGATTCAA 19034
 QY 2236 AGGACCTGAATAGCCAGATGTTCCAGATACAAATGAAAGTGAACCCAGCACAGCAGATA 2295
 Db 19033 AGGACCTGAATAGCCAGATGTTCCAGATACAAATGAAAGTGAACCCAGCACAGCAGATA 18974
 QY 2296 CCGTTGTTCCAACTGCTCTATAGATTTACAGCTGTATCCACAGGTTAGCTTCAACT 2355
 Db 18973 CCGTTGTTCCAACTGCTCTATAGATTTACAGCTGTATCCACAGGTTAGCTTCAACT 18914
 QY 2356 CTAATATCAATGATATAGAAAGTAATCTGAGAGTGTATCTGTTTCGTAATGCAATG 2415

Db 18913 CTAATATCAATGATATAGAAAGTAATCTGAAGGTGGATCTAGTTTCGTAATGCAATG 18854
 QY 2416 AGATTCCTGACTGAAACACTTCGAAAGAGGCTTGGTTTGGCCGCAAGACAGCCTA 2475
 Db 18853 AGATTCCTGACTGAAACACTTCGAAAGAGGCTTGGTTTGGCCGCAAGACAGCCTA 18794
 QY 2476 CTTGGGTTCTCTGATTCAGAAAGCTCCAAACTGTATGAACTGCACAGTCAAAATTTACTTTA 2535
 Db 18793 CTTGGGTTCTCTGATTCAGAAAGCTCCAAACTGTATGAACTGCACAGTCAAAATTTACTTTA 18734
 QY 2536 CCAACGGGACACCATTCGCGAGCATGTGGAAAGTATTTTGT 2579
 Db 18733 CCAACGGGACACCATTCGCGAGCATGTGGAAAGTATTTTGT 18690
 RESULT 12
 AC008771 123169 bp DNA linear PRI 28-FEB-2001
 LOCUS Homo sapiens chromosome 5 clone CTD-2015H6, complete sequence.
 DEFINITION AC008771
 ACCESSION AC008771
 VERSION AC008771.4 GI:13162500
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 123169)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 COMMENT On Feb 28, 2001 this sequence version replaced gi:7709299.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www-shgc.stanford.edu
 Quality: Phrap Quality >=40 99.3% of Sequence;
 Estimated Total Number of Errors is 0.9.
 STS Content:
 SHGC-84992 GS3847
 WI-15051 G23648
 WI-17790 G24283.
 FEATURES
 Location/Qualifiers
 source
 1..123169
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2015H6"

ORIGIN

Query Match 34.0%; Score 2252.8; DB 9; Length 123169;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2257; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 316 CAGATGACACAGATTTATCTCCAGATGTACAAATGCAATATGATCTTAACCACTGCTCAG 375
 Db 43251 CAGATGACACAGATTTATCTCCAGATGTACAAATGCAATATGATCTTAACCACTGCTCAG 43310
 QY 376 TTTCTTCAGAGTTGGCTTCCTCACAGCGAACTTCACTGCTCCCAAGACCAAGAGTSCG 435
 Db 43311 TTTCTTCAGAGTTGGCTTCCTCACAGCGAACTTCACTGCTCCCAAGACCAAGAGTSCG 43370

Qy	436	TTAATAGTTGTGCTCATCAGAAACAAGCTATGGAAACAAATAGAGAGTTCCCTGAATGAAA	495
Db	43371	TTAATAGTTGTGCTCATCAGAAACAAGCTATGGAAACAAATAGAGAGTTCCCTGAATGAAA	43430
Qy	496	AAACACCTCAAGGACCTTACTCTATACAAAATGAAAATAATCTACACGACTTGTATCTTC	555
Db	43431	AAACACCTCAAGGACCTTACTCTATACAAAATGAAAATAATCTACACGACTTGTATCTTC	43490
Qy	556	TTTCTTCTGTGGATGGTGTACTTCAGATGAAATCCAGCGTTATATATGGACGATGTA	615
Db	43491	TTTCTTCTGTGGATGGTGTACTTCAGATGAAATCCAGCGTTATATATGGACGATGTA	43550
Qy	616	GTAACCTATCTGTGATCTGATAAGTGACATGGGTAACTTAGTTCATCGAACCCATAGTG	675
Db	43551	GTAACCTATCTGTGATCTGATAAGTGACATGGGTAACTTAGTTCATCGAACCCATAGTG	43610
Qy	676	AAGAAGATATTAAAAATTTATTGCCAGATGATTTTAAAGTCTAATCGAGATTCCTTGATTG	735
Db	43611	AAGAAGATATTAAAAATTTATTGCCAGATGATTTTAAAGTCTAATCGAGATTCCTTGATTG	43670
Qy	736	GATTGGATTTATCTTCAGTGTGAGATACCTCCCTGTGTTTCTTCAACAGACCATGATGTG	795
Db	43671	GATTGGATTTATCTTCAGTGTGAGATACCTCCCTGTGTTTCTTCAACAGACCATGATGTG	43730
Qy	796	ATACTGTGAGAAACAACAGATGATCAGTCTCTGAATTACAAATAGAGAAATCGGAG	855
Db	43731	ATACTGTGAGAAACAACAGATGATCAGTCTCTGAATTACAAATAGAGAAATCGGAG	43790
Qy	856	GAATCAAGAATTTGGTATAAAGTAGATACAACTTTTCAGATTCCTTATTAATACAGTG	915
Db	43791	GAATCAAGAATTTGGTATAAAGTAGATACAACTTTTCAGATTCCTTATTAATACAGTG	43850
Qy	916	GAACAGAAATTTAAAGATATAAAGATCTTTAAATCAGTTAGAACTAAATGTTGATTTTA	975
Db	43851	GAACAGAAATTTAAAGATATAAAGATCTTTAAATCAGTTAGAACTAAATGTTGATTTTA	43910
Qy	976	ACATGTCACTGTCTTTGACTCGCAAAAGTTCCAAATGTTTCATGCCAAAGCAAGCTAC	1035
Db	43911	ACATGTCACTGTCTTTGACTCGCAAAAGTTCCAAATGTTTCATGCCAAAGCAAGCTAC	43970
Qy	1036	AACACAGAGCCAGCCATGTGGATTAATAAGATCTTGGCTTACTTAAAGAGGAAGTAG	1095
Db	43971	AACACAGAGCCAGCCATGTGGATTAATAAGATCTTGGCTTACTTAAAGAGGAAGTAG	44030
Qy	1096	ATGTGGCAGTCAAACTGCGCGCAGAAATGTTTAAAGAGAGGCGCAAGCAAGTGTCTTGA	1155
Db	44031	ATGTGGCAGTCAAACTGCGCGCAGAAATGTTTAAAGAGAGGCGCAAGCAAGTGTCTTGA	44090
Qy	1156	CCTGCGACCTTCGAAAAATGAAGATTTATGCTTAAATGATTTCAAATTCAGAGATGAAA	1215
Db	44091	CCTGCGACCTTCGAAAAATGAAGATTTATGCTTAAATGATTTCAAATTCAGAGATGAAA	44150
Qy	1216	ATTTCAAATTAACCTGACTTTTCCTTTACGAAAGATAAGACTGTTATAAAAAATCTGCAC	1275
Db	44151	ATTTCAAATTAACCTGACTTTTCCTTTACGAAAGATAAGACTGTTATAAAAAATCTGCAC	44210
Qy	1276	AAGAAGCTCAAAAAGTTTAGACCTTTAAGATTAATGATTAATCCAAGATTCCTCTTCAG	1335
Db	44211	AAGAAGCTCAAAAAGTTTAGACCTTTAAGATTAATGATTAATCCAAGATTCCTCTTCAG	44270
Qy	1336	CTTTACATGTTTCCAGTAAAGATGTCCGCTCCTCATGTGCTCTCTCGCTCTGGGT	1395
Db	44271	CTTTACATGTTTCCAGTAAAGATGTCCGCTCCTCATGTGCTCTCTCGCTCTGGGT	44330
Qy	1396	CTATGTGTGGATCATTAATTGAAAGTAAAGACCGGGTGATTTTTTACTCAGCATGAAC	1455
Db	44331	CTATGTGTGGATCATTAATTGAAAGTAAAGACCGGGTGATTTTTTACTCAGCATGAAC	44390
Qy	1456	ATAAAGATATATACAGATCGAGTCACTATACATGAAGAATACAGAACAGTGTGTTTC	1515
Db	44391	ATAAAGATATATACAGATCGAGTCACTATACATGAAGAATACAGAACAGTGTGTTTC	44450
Qy	1516	TAGTGGGGAACCATTTCAAGAGAATGATCTTTTGAACAGAGAAAATGTAAACACATAC	1575

Db	44451	TAGGTGGGAACCA	TTTCAAGAGAGATGATCTTTTGAACAGGAAANA	TGTA	AAAGCATAC	44510
Qy	1576	TCCTTCAGTCA	TTAAATTGAAGGGATGG	AAAGACAGAAAGATAGATCCTGAC	CAGTAA	1635
Db	44511	TCCTTCAGTCA	TTAAATTGAAGGGATGG	AAAGACAGAAAGATAGATCCTGAC	CAGTAA	44570
Qy	1636	TCAGAGCTGAGTCTTTGGATG	TGTTGGTGCACACAGTCTTACAGTCTG	TGAGATCTCAAGAGG	1695	
Db	44571	TCAGAGCTGAGTCTTTGGATG	TGTTGGTGCACACAGTCTTACAGTCTG	TGAGATCTCAAGAGG	44630	
Qy	1696	GGCTTTCTGGCACTCATG	TCCAGAGTCTTCTGATTTGTGAAGGTTTTATTA	TACTTT	1755	
Db	44631	GGCTTTCTGGCACTCATG	TCCAGAGTCTTCTGATTTGTGAAGGTTTTATTA	TACTTT	44690	
Qy	1756	TTTCAAGCAATGATATGGAT	TGGCAAGACTTAGATTA	CTTTTAATTTGATG	AGGCGCAA	1815
Db	44691	TTTCAAGCAATGATATGGAT	TGGCAAGACTTAGATTA	CTTTTAATTTGATG	AGGCGCGCAA	44750
Qy	1816	AAAGTGCCCACTAA	TTAGTGCTGAACTTTGATGCCCTTTCTGACAGAA	CAGTAGTCTTC	1875	
Db	44751	AAAGTGCCCACTAA	TTAGTGCTGAACTTTGATGCCCTTTCTGACAGAA	CAGTAGTCTTC	44810	
Qy	1876	AGACCACTAACTAA	AGTCTTTTGAAGAAATGTAAATGACTCTAAATCG	GAATGAATC	1935	
Db	44811	AGACCACTAACTAA	AGTCTTTTGAAGAAATGTAAATGACTCTAAATCG	GAATGAATC	44870	
Qy	1936	AGATAGATATGAAGGCTTAGATG	TGGAACATCAATATATATTTCAATGCAGAG	1995		
Db	44871	AGATAGATATGAAGGCTTAGATG	TGGAACATCAATATATATTTCAATGCAGAG	44930		
Qy	1996	CAGGAGCTATTGGGGAAGT	CATGGTATTAATATATTTGTGAACAGTTGATAA	CAAAA	2055	
Db	44931	CAGGAGCTATTGGGGAAGT	CATGGTATTAATATATTTGTGAACAGTTGATAA	CAAAA	44990	
Qy	2056	ATCAATAGAAAAATG	CGCTTTCTTTAGGAGAAAAAGCACTATCCAGTTCA	CAAGGCT	2115	
Db	44991	ATCAATAGAAAAATG	CGCTTTCTTTAGGAGAAAAAGCACTATCCAGTTCA	CAAGGCT	45050	
Qy	2116	TACCTACCACTAAGTCTG	AGATTACAAATCAATATCAGTCTCTGATATTA	CAGTCAAT	2175	
Db	45051	TACCTACCACTAAGTCTG	AGATTACAAATCAATATCAGTCTCTGATATTA	CAGTCAAT	45110	
Qy	2176	CTGTGTGAGGGCCAGACCT	TAGCAATGTTTAGCCTCCATCAAGAACAGGATTC	CAA	2235	
Db	45111	CTGTGTGAGGGCCAGACCT	TAGCAATGTTTAGCCTCCATCAAGAACAGGATTC	CAA	45170	
Qy	2236	AGGACCTGTAATAGCCAGAT	TTTCCAGATACAATGAAAGTGAA	CCCAAGCACA	2295	
Db	45171	AGGACCTGTAATAGCCAGAT	TTTCCAGATACAATGAAAGTGAA	CCCAAGCACA	45230	
Qy	2296	CCGTGTGTCAAATCACT	TGTGCTATAGATCTACAGCTGATCCACAGGTTAG	CTTCAACT	2355	
Db	45231	CCGTGTGTCAAATCACT	TGTGCTATAGATCTACAGCTGATCCACAGGTTAG	CTTCAACT	45290	
Qy	2356	CTAATTA	CAATGATAGAAAGTAAATCTGAAGGTTG	GAATCTAGTTTCGTAAC	2415	
Db	45291	CTAATTA	CAATGATAGAAAGTAAATCTGAAGGTTG	GAATCTAGTTTCGTAAC	45350	
Qy	2416	AAGATTCTGTACCTG	AAAAACACTTGC	AAAGAGGCTTGTGTTTGGCCAGAAC	2475	
Db	45351	AAGATTCTGTACCTG	AAAAACACTTGC	AAAGAGGCTTGTGTTTGGCCAGAAC	45410	
Qy	2476	CTTGGGTTCTCTGAT	TGAGAGCTCCAACTGTATGA	CTGAACTGCCAAGTCA	2535	
Db	45411	CTTGGGTTCTCTGAT	TGAGAGCTCCAACTGTATGA	CTGAACTGCCAAGTCA	45470	
Qy	2536	CCAAACGGGACACCA	TTGCGGAGCATGTGGG	AAAGTATTTTGT	2579	
Db	45471	CCAAACGGGACACCA	TTGCGGAGCATGTGGG	AAAGTATTTTGT	45514	

AC018764/c
 LOCUS AC018764 126052 bp DNA linear PRI 13-MAR-2003
 DEFINITION Homo sapiens chromosome 5 clone CTD-2327L5, complete sequence.
 ACCESSION AC018764
 VERSION AC018764.8 GI:28933541
 HTG.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Unpublished
 2 (bases 1 to 126052)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (19-DEC-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (09-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 5 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (30-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 6 (bases 1 to 126052)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 Direct Submission
 Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Mar 13, 2003 this sequence version replaced gi:28144392.
 Draft Sequence produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.sbgc.stanford.edu
 Quality: Phrap Quality >=40 99.6% of Sequence;
 Estimated total Number of Errors is 0.6.
 NOTE: Bacterial transposon excised at 89327.
 Location/Qualifiers
 1..126052
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2327L5"

FEATURES
 source
 Query Match 28.8%; Score 1910.4; DB 9; Length 126052;
 Best Local Similarity 99.9%; Pred. No. 2e-298;
 Matches 1922; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

ORIGIN
 4709 GGTGAATTCAGGAGGATCTGAAGCCAACTTCGCTCAGCAATATCTAAATGATCT 4768
 DB 97844 GGTGAATTCAGGAGGATCTGAAGCCAACTTCGCTCAGCAATATCTAAATGATCT 97785
 4769 TGATAGTCTCTGATACCTGATGCCATGGGAGCCTCCAACTCTAGTTTACCATTAGA 4828
 DB 97784 TGATAGTCTCTGATACCTGATGCCATGGGAGCCTCCAACTCTAGTTTACCATTAGA 97725
 4829 AATAGAATTAGTGTCTTTCATTATAGAACATCTTTTCTAGTGAAGAAATCTGCCATATTA 4888
 DB 97724 AATAGAATTAGTGTCTTTCATTATAGAACATCTTTTCTAGTGAAGAAATCTGCCATATTA 97665
 4889 CATATTGCAACTAATTTGTTTAAACTAACTCCAGCACTAAAGCTGAAGTGCACCAACA 4948

Db 97664 CATATTGCAACCTAATTTGTTTAAAACTAACTCCAGCACTAAAGCTGAATGCCCAACA 97605
 QY 4949 CTAAAGATATAAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGAGGAAT 5008
 Db 97604 CTAAAGATATAAATATGCTGATTTTGAACACACATAAGCTTTGCTCTTTTAGGAGGAAT 97545
 QY 5009 GATCTTTTCAATCAATTAGCAATATTTAAATATCTAAAAATTTAAGAGATCCATCTT 5068
 Db 97544 GATCTTTTCAATCAATTAGCAATATTTAAATATCTAAAAATTTAAGAGATCCATCTT 97485
 QY 5069 TCTGTAGCTTTACAAATTAATTTAAGTACTAAAGACACAGGATTTCTTTTAAAGAAATTTA 5128
 Db 97484 TCTGTAGCTTTACAAATTAATTTAAGTACTAAAGACACAGGATTTCTTTTAAAGAAATTTA 97425
 QY 5129 TAGCAATTTACTGTGTATTTAAATCTAAGCCAAAGTATCTGCATTAAGGTATACCTCTT 5188
 Db 97424 TAGCAATTTACTGTGTATTTAAATCTAAGCCAAAGTATCTGCATTAAGGTATACCTCTT 97365
 QY 5189 TATGCCAATATGATTTTAAAGAGGCTTTTTCAGATGTAACCTTATGAAGGAAATATC 5248
 Db 97364 TATGCCAATATGATTTTAAAGAGGCTTTTTCAGATGTAACCTTATGAAGGAAATATC 97305
 QY 5249 TGCCTTGTGTATATGCCAGTTAGATCTGGTTTCTAAAGCTGTCAAAATTTGTAATTCAG 5308
 Db 97304 TGCCTTGTGTATATGCCAGTTAGATCTGGTTTCTAAAGCTGTCAAAATTTGTAATTCAG 97245
 QY 5309 TGGCAAAAAACCGATTTTGAGGCTTTAGACTTATAATCTTTGAATAAAACTGATAACT 5368
 Db 97244 TGGCAAAAAACCGATTTTGAGGCTTTAGACTTATAATCTTTGAATAAAACTGATAACT 97185
 QY 5369 TATTTGTATAATTTGAGTGGAGACCTCTCCATAATTAGATAAATCTTTTGGATTAT 5428
 Db 97184 TATTTGTATAATTTGAGTGGAGACCTCTCCATAATTAGATAAATCTTTTGGATTAT 97125
 QY 5429 AATCAGAAATTTTGCCTTTTCTTCTCAATATTTACATATGTATGTATATATATCCAC 5488
 Db 97124 AATCAGAAATTTTGCCTTTTCTTCTCAATATTTACATATGTATGTATATATATCCAC 97065
 QY 5489 ATATATAGTTTCCCTGATTAATGATTAATTAATTAATTTCCGGTGGCTTCAGGACTTTT 5548
 Db 97064 ATATATAGTTTCCCTGATTAATGATTAATTAATTAATTTCCGGTGGCTTCAGGACTTTT 97005
 QY 5549 TCGTCTATATTTAAGTATATTTGTTTTATAGCAAGAACATATCTGAATGTTTTATAAA 5608
 Db 97004 TCGTCTATATTTAAGTATATTTGTTTTATAGCAAGAACATATCTGAATGTTTTATAAA 96945
 QY 5609 TCTTTAATAATTTATGTAGTAATATTTTGTATGCAATGCAATATTTTTCCTC 5668
 Db 96944 TCTTTAATAATTTATGTAGTAATATTTTGTATGCAATGCAATATTTTTCCTC 96886
 QY 5669 CTTTCCCTCCAACTATACCTGATTTTCCACTCTAAGAGTACTGACGACGGGCA 5728
 Db 96885 CTTTCCCTCCAACTATACCTGATTTTCCACTCTAAGAGTACTGACGACGGGCA 96826
 QY 5729 GATGACCTTGAAGTACTGATTTAGCAATAAATGAAGCCTGAAACAGGTTTTTTTAC 5788
 Db 96825 GATGACCTTGAAGTACTGATTTAGCAATAAATGAAGCCTGAAACAGGTTTTTTTAC 96766
 QY 5789 TTCCACTTTAATCCCTAGAAATTTCTTGGCACTTCGCATATTTTTCATTGACACCTGGTGT 5848
 Db 96765 TTCCACTTTAATCCCTAGAAATTTCTTGGCACTTCGCATATTTTTCATTGACACCTGGTGT 96706
 QY 5849 ATAAGTAAATTTAAATGAACATAATTTTGCATATTTTAAATCTTTTATATGGTAG 5908
 Db 96705 ATAAGTAAATTTAAATGAACATAATTTTGCATATTTTAAATCTTTTATATGGTAG 96646
 QY 5909 TTATTTTATACAGGATATTAACATAGTAATTAATCTTATGTTTCAATTTCTTACAG 5968
 Db 96645 TTATTTTATACAGGATATTAACATAGTAATTAATCTTATGTTTCAATTTCTTACAG 96586
 QY 5969 ACCTTTCTCTTTTACTTCAAAACAGCAAAAAAGTGGGGGCATATTTGTAGTCTCTGCAATTT 6028

Best Local Similarity 98.4%; Pred. No. 3.5e-286;
Matches 1896; Conservative 0; Mismatches 20;

	Best Local Similarity	98.4%; Pred. No. 3.5e-286;	Matches 1896; Conservative 0; Mismatches 20; Indels 10; Gaps 4;
Qy	4709	GGITGAAATTTTCAGGCAGCATCTGAAGCGCAACTTCTGGCTCAGCAATTATCTAATAATGATCT	4768
Dd	626	GGITGAAATTTTCAGGCAGCATCTGAAGCGCAACTTCTGGCTCAGCAATTATCTAATAATGATCT	685
Qy	4769	TGATAGTGCTCTGATACCTGTGATCCATGTGGGACCTCCAACHTCTAGTTTTACCATTAGA	4828
Dd	686	TGATAGTGCTCTGATACCTGTGATCCATGTGGGACCTCCAACHTCTAGTTTTACCATTAGA	745
Qy	4829	AATAGAATTAGTGTGTTCATTATAGAACAPCTTTTTTAGTGAAGAATGTGCCATATTA	4888
Dd	746	AATAGAATTAGTGTGTTCATTATAGAACATCTTTTTTAGTGAAGAATGTGCCATATTA	805
Qy	4889	CATATTCGAACCTAATTTGTTTAAAACTPAACCTCCAGCACCTAAAGCTGAAAATGCCCAAACA	4948
Dd	806	CATATTCGAACCTAATTTGTTTAAAACTPAACCTCCAGCACCTAAAGCTGAAAATGCCCAAACA	865
Qy	4949	CTAAAACTAATAATCTGTGATTTTGAACACATAGCTTTCCTCTTTAGGCAGCAAT	5008
Dd	866	CTAAPAGTATAA--ATGTCTGATTTTTGAACAACATAGCTTTCCTCTTTAGGCAGCAAT	923
Qy	5009	GATCTTTTCAAAATCATTAGCACATAATTAAATATCTAAAAATTTAAGAGATCCATACTT	5068
Dd	924	GATCTTTTCAAAATCATTAGCACATAATTAAATATCTAAAAATTTAAGAGATCCATACTT	983
Qy	5069	TCGTAGCTTTACAATTAATTAAAGTACTAAAAAGACAGGAATTCCTTTAAGHAATTTA	5128
Dd	984	TCGAAGCTTTACAATTAATTAAAGTACTAAAAAGACAGGAATTCCTTTAAGHAATTTG	1043
Qy	5129	TAGCAATTAAGTGTGTATTTTAAATGCTTAAGCCAAAGTATCTGCACTTAGGTATACCTCTT	5188
Dd	1044	TAGCAATTAAGTGTGTATTTTAAATGCTTAAGCCAAAGTATCTGCACTTAGGTATACCTCTT	1103
Qy	5189	TATGCCAATAATGAATTTAATGAAGGTCTTTTTCAGATGTAACCTTAAGAAGAAATPATC	5248
Dd	1104	TATGCCAATAATGAATTTAATGAAGGTCTTTTTCAGATGTAACCTTAAGAAGAAATPATC	1163
Qy	5249	TGCTTTCTGTATATGCCAGTTAGAAATACTGGTTTCTAAAGCTCTGTCAAATTTGATTTTCAG	5308
Dd	1164	TGCTTTCTGTATATGCCAGTTAGAAATACTGGTTTCTAAAGCTCTGTCAAATTTGATTTTCAG	1223
Qy	5309	TGGCACA AAA AAC CAG TTTT TG AG GT CT TT A GA CT T TA T AA T CT TTT GA AT P AA AC CT G ATA ACT	5368
Dd	1224	TGGCACA AAA AAC CAG TTTT TC G AG GT CT TT A GA CT T TA T AA T CT TTT GA AT P AA AC CT G ATA ACT	1283
Qy	5369	TATTTGATTAATTTGGAGTGGAGACCTACCTCCATTAATTTAGATAAACTCTTTTGGATTAT	5428
Dd	1284	TATTTGATTAATTTGGAGTGGAGACCTACCTCCATTAATTTAGATAAACTCTTTTGGATTAT	1343
Qy	5429	AACTCAGAAATTTTGCTTTTTTCTCTCAAATTTATACATATGTATGTATTATATATCCAT	5488
Dd	1344	AACTCAGAAATTTTGCTTTTTTCTCTCAAATTTATACATATGTATGTATTATATATCCAT	1403
Qy	5489	ATATATAGTTTTCCCTGATTAATTTGATNTATAAATTTTGGGGTGCCTTCAGCACTTTT	5548
Dd	1404	ATATATAGTTTTTCCCTGATTAATTTGATNTATAAATTTTGGGGTGCCTTCAGCACTTTT	1463
Qy	5549	TGCTTCTPATAATTTAAAGTATATGTTTTTATAGCAAGAACAATATCTCTGAA--TGTTTTTATA	5606
Dd	1464	TGCTTCTPATAATTTAAAGTATATGTTTTTATAGCAAGAACAATATCTCTGAAATGCTGTTTATA	1523
Qy	5607	AATCTTTTAATAATTTATATAGTATAATTTTTTGATATCAAAATPGCAATATTTTTTTTCC	5666
Dd	1524	AATCTTTTAATAATTTATATAGTATAATTTTTTGATATCAAAATPGCAATATTTTTTTTCC	1582
Qy	5667	TCCTTTTCCCTCCAAAACATATACCTGATTTTACCACTTCTTAAGAGTGACGTGACGAGGGC	5726
Dd	1583	TCCTTTTCCCTCCAAAAGTATACCACTGATTTTACCACTTCTTAAGAGTGACGTGACGAGGGC	1642
Qy	5727	CAGATGACCCCTTGAAGTAGTCTATTATGTAGCAATAAATGAAGGCTCGAAACAGGTTTTTTT	5786

1543	Db	CAGATGACCCCTTGAAGTAGTGCATATATGTAGCAATAAATGAAGCCTGNAACAGAGTTTTTT	1702
5787	Qy	ACTTCCACTTTAATCCCTTAGAAATTTCTTGGCAACTTCGCGAATATTTTCATGTACACTGGT	5846
1703	Db	ACTTCCACTTTAATCCCTTAGAAATTTCTTGGCAACTTCGCGAATATTTTCATGTACAGCAAT	1762
5847	Qy	GTATAAGTATAAATTTAAATGAACATAATTAATCTTTTGCATATATTTAAATCTTTTATATGGT	5906
1763	Db	GTATAAGTATAAATTTAAATGAACATAATTAATCTTTTGCATATATTTAAATCTTTTATATGGT	1822
5907	Qy	AGTTATTTTTTATAACAGGATATTTAAACATAAGTTTAAATCCTATGTATTTGAAATTCGTTAC	5966
1823	Db	AGTTATTTTTTATAACAGGATATTTAAACATAAGTTTAAATCCTATGTATTTGAAATTCGTTAC	1882
5967	Qy	AGAGCTTTCCCTTTTACTTCAACACAGAAAAGTGGGGGCGATATGTAGTCCCTGTCAT	6026
1883	Db	AGAGCTTTCCCTTTTACTTCAAAATGCAAAAAGTGGGGGCGATATGTAGTCTTGTGCAT	1942
6027	Qy	TTAAGTTATGTAAAAAATTTAATCATTTATTTTCATGCTTTTAAACATTTCTCATGTGTAATA	6086
1943	Db	TTAAGTTATGTAAAAAATTTAATCATTTATTTTCATGCTTTTAAACATTTCTCATGTGTAATA	2002
6087	Qy	TATGTTTTTGTATCAAAAACACATCATATATTTCAAGAAAAGAAATATGTTHAATAGCC	6146
2003	Db	TATGTTTTTGTATCAAAAACACATCATATATTTCAAGAAAAGAAATATGTTHAATAGCC	2062
6147	Qy	CTGTTTTAAAGAAAATTTATTAAGAGCATCTCAACTTTGAAGATCAAGTCAAGGTTATAAC	6206
2063	Db	CTGTTTTAAAGAAAATTTATTAAGAGCATCTCAACTTTGAAGATCAAGTCAAGGTTATAAC	2122
6207	Qy	TCAGGATCTGAGTCTCAAGCTAGGAGAGACTGAGAAATTTTAATCAGTTTGGGCATATAG	6266
2123	Db	TCAGGATCTGAGTCTCAAGCTAGGAGAGACTGAGAAATTTTAATCAGTTTGGGCATATAG	2182
6267	Qy	TTTGGACTGAATCACATCTGTAGTACTTAGCCAAACAGCAATTTGGAGGAGAAATATCAGCC	6326
2183	Db	TTTGGACTGAATCACATCTGTAGTACTTAGCCAAACAGCAATTTGGAGGAGAAATATCAGCC	2242
6327	Qy	TTCTGGAAGTAGTACTTCTGAAACAATGTAAAGTGTGCGAGATATTCAAATAAAATGGCA	6386
2243	Db	TTCTGGAAGTAGTACTTCTGAAACAATGTAAAGTGTGCGAGATATTCAAATAAAATGGCA	2297
6387	Qy	ACCTGTTATTAATTTGTGAAATTTATTTGAAATGTTGTAAGATGAAPCAATTCGCATATCAA	6446
2298	Db	ACCTGTTATTAATTTGTGAAATTTATTTGAAATGTTGTAAGATGAAPCAATTCGCATATCAA	2357
6447	Qy	ACCCAAATTTATGTTTTCTAAATATAGTGTATGTATTTCTGCCATGTAAAGTAATGGAACAGT	6506
2358	Db	ACCCAAATTTATGTTTTCTAAATATAGTGTATGTATTTCTGCCATGTAAAGTAATGGAACAGT	2417
6507	Qy	CTTAAATTAACCAAAATGGTAGGGCTGTTCATGATGGGACAGCTTTCGATTTGTTTTC	6566
2418	Db	CTTAAATTAACCAAAATGGTAGAGAGCTGTTCATGATGGGATAGCTTTGGAATTTGTTTTC	2477
6567	Qy	ATAAAATCTCTACATTCAAATAAAAATTTGAAATTTATGTGCTGAAAGTTTGGAGGCAATTT	6626
2478	Db	ATAAAATCTCTACATTCAAATAAAAATTTGNAATTTATGTGCTGAAAGTTTGGAGGCAATTT	2537
6627	Qy	TGAAGT	6632
2538	Db	TGAAAT	2543

RESULT 15	AK090896	2622 bp	mrna	linear	PRI 15-JUL-2002
LOCUS	AK090896				
DEFINITION	Homo sapiens cDNA FLJ33577 fis, clone BRAWY2011480.				
ACCESSION	AK090896				
VERSION	AK090896.1 GI:21749142				
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	QY	5189	TATGCCAATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAATATC	5248
AUTHORS	1 Tashiro, H., Yanaizaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, I., Sato, H., Wakanatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuko, Y., Nagai, K. and Isogai, T.	Db	1420	TATGCCAATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAATATC	1479
TITLE	Unpublished	QY	5249	TGCTTTTGATATATCCAGTTAGAACTACTGGTTCTTAAAGTCTGTCAAAATTTGATTTT	5308
JOURNAL	2 (bases 1 to 2622)	Db	1480	TGCTTTTGATATATCCAGTTAGAACTACTGGTTCTTAAAGTCTGTCAAAATTTGATTTT	1539
AUTHORS	Isogai, T. and Yamamoto, J.	QY	5309	TGGCAAAAAACCACTTTTGGAGTCTTACATTAATAATCTTTGAATAAACTGATAACT	5369
TITLE	Direct Submission	Db	1540	TGGCAAAAAACCACTTTTGGAGTCTTACATTAATAATCTTTGAATAAACTGATAACT	1599
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7	QY	5369	TATTTGTATAATGGAGTGGAGACCTACTCCATAATTAGATAAACTCTTTTGGATAT	5428
COMMENT	Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@ri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' - end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	Db	1600	TATTTGTATAATGGAGTGGAGACCTACTCCATAATTAGATAAACTCTTTTGGATAT	1659
FEATURES	Location/Qualifiers	QY	5429	AATCAGAAATTTGCCCTTTTCTCTCTCAAAATTTATACATATGTATATATATATAC	5488
source	1. 2622	Db	1660	AATCAGAAATTTGCCCTTTTCTCTCTCAAAATTTATACATATGTATATATATATAC	1719
	/organism="Homo sapiens"	QY	5489	ATATATAGTTTCCCTGATTAATCGATATTAATAATTTGGGGTCTTCAGGACTTTT	5548
	/mol_type="mRNA"	Db	1720	ATATATAGTTTCCCTGATTAATCGATATTAATAATTTGGGGTCTTCAGGACTTTT	1779
	/db_xref="taxon:9606"	QY	5549	TGCTTCTATATTAAGTATATTTTATAGCAAGAACATATCTGAATGTTTATATAA	5608
	/clone="BRAMV2011480"	Db	1780	TGCTTCTATATTAAGTATATTTTATAGCAAGAACATATCTGAATGTTTATATAA	1839
	/tissue_type="amygdala"	QY	5609	TCTTTAATAATTTATATGATAGTAATATTTTGTATATCAATGATATATTTTTCCTC	5668
	/clone_lib="BRAMY2"	Db	1840	TCTTTAATAATTTATATGATAGTAATATTTTGTATATCAATGATATATTTTTCCTC	1898
	/note="cloning vector: pME18SFL3"	QY	5669	CTTTCCTTCCAAACATATACCTGTATTTACCACTTCTAAGAGTCACTGACACGGGCA	5728
ORIGIN		Db	1899	CTTTCCTTCCAAACATATACCTGTATTTACCACTTCTAAGAGTCACTGACACGGGCA	1958
	Query Match 25.2%; Score 1668.8; DB 9; Length 2622;	QY	5729	GATGACCTTGAAGTAGTCATTATGTAGCAATTAATGAAGCTGAAACAGGTTTTTTTAC	5788
	Best Local Similarity 99.8%; Pred. No. 4.8e-259; Mismatches 1; Gaps 1;	Db	1959	GATGACCTTGAAGTAGTCATTATGTAGCAATTAATGAAGCTGAAACAGGTTTTTTTAC	2018
	Matches 1681; Conservative 0; Indels 2;	QY	5789	TTCCACTTTAACTCTTAGAAATTTCTTGGCACTTGGCATATTTTCATTGACACATGTTG	5848
		Db	2019	TTCCACTTTAACTCTTAGAAATTTCTTGGCACTTGGCATATTTTCATTGACACATGTTG	2078
		QY	5849	ATAAGTATAAATTTAAATGAACATAATCTTTTGTGATATTTTAAATTTCTTTATATG	5908
		Db	2079	ATAAGTATAAATTTAAATGAACATAATCTTTTGTGATATTTTAAATTTCTTTATATG	2138
		QY	5909	TTATTTTATTAACAGGATATTAACATAAGTTTAAATCTCTATATTTGAAATTTGTTAC	5968
		Db	2139	TTATTTTATTAACAGGATATTAACATAAGTTTAAATCTCTATATTTGAAATTTGTTAC	2198
		QY	5969	AGCTTTTCTCTTTACTTTCAACAGCAAAAAAGTGGGGGCATATTTGAGTCTCTGTCAT	6028
		Db	2199	AGCTTTTCTCTTTACTTTCAACAGCAAAAAAGTGGGGGCATATTTGAGTCTCTGTCAT	2258
		QY	6029	AAGTTATGTAAAAAATTTAATCATTTATTTGATGTTTAAACATTTCTCATGTGTAATA	6088
		Db	2259	AAGTTATGTAAAAAATTTAATCATTTATTTGATGTTTAAACATTTCTCATGTGTAATA	2318
		QY	6089	TGTTTTTGTATCAAAACACTCATATATTTCAAGAAAGAAATTTATGTTAAATAGCCCT	6148
		Db	2319	TGTTTTTGTATCAAAACACTCATATATTTCAAGAAAGAAATTTATGTTAAATAGCCCT	2378
		QY	6149	GTTTTAAAGAAAAATTTTATGAAGCATCTCAACTTTGAAGATCAAGTCAAAGTTATACT	6208
		Db	2379	GTTTTAAAGAAAAATTTTATGAAGCATCTCAACTTTGAAGATCAAGTCAAAGTTATACT	2438
		QY	6209	AGGATCTGAGGCTCAAGCTAGGAGACTGAGAAATTTTAAATCAGTTTGGGCATATAGT	6268
		Db	2439	AGGATCTGAGGCTCAAGCTAGGAGACTGAGAAATTTTAAATCAGTTTGGGCATATAGT	2498
		QY	6269	TGGACTGAATCAATCTGTAGTACTTGTAGCCAAAAGACAATTTGGAGGAGAAATATCAG	6328

Db	2499	TCGACTGAATCACAATCTGTAGTACTTAGCCAAAGACAAATTTGGAGGAGAAATTCAGCCTT	2558
Qy	6329	CTGGAAGTAGCTACTTCCTGAAACAATGTAAAGTGTCCGAGATATTCATAATAAATGGCAAC	6388
Db	2559	CTGGAAGTAGCTACTTCCTGAAACAATGTAAAGTGTCCGAGATATTCATAATAAATGGCAAC	2618
Qy	6389	CTGT	6392
Db	2619	CTGT	2622

Search completed: May 18, 2004, 02:10:37
Job time : 16485.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:46:50 ; Search time 1526.9 Seconds
(without alignments)
18451.753 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 actccggcgccgggtagctc.....ttggaggccattttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 1 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2000s:*
5: Geneseqn2000s:*
6: Geneseqn2000s:*
7: Geneseqn2000s:*
8: Geneseqn2000s:*
9: Geneseqn2000s:*
10: Geneseqn2000s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6632	100.0	6632	3	Aaz50069 Human Sma
2	5290	79.8	5402	6	Abk11709 DNA encod
3	1514	22.8	1572	3	Aac98041 Human col
4	850	12.8	969	6	AbL89832 Human pol
5	665.4	10.0	693	3	Aaa44354 Human col
6	580.2	8.7	701	6	Abq57538 Human col
7	560.6	8.5	4839	3	Aaz50068 Human Sma
8	556.2	8.4	1977	9	Aac64443 Human SAR
9	500	7.5	4823	3	Aaz50070 Xenopus S
10	472.4	7.1	475	8	Ach15512 Human adu
11	444	6.7	455	5	Aah81649 Human dif
12	389.4	5.9	391	8	Ach20104 Human adu
13	286	4.3	317	4	Aah35041 Human col
14	199.6	3.0	393	7	Abx48202 Bovine ES
15	191.8	2.9	4068	4	AbL28217 Drosophil
16	187.4	2.8	277	2	Aat23367 Human gen
17	181.6	2.7	289	2	Aav88493 EST clone
18	180.8	2.7	6784	4	AbL28216 Drosophil
19	172	2.6	556	4	Aak11740 Human bra
20	168	2.5	168	4	Aak24333 Human bra
21	124	1.9	387	6	Abn93507 Gene #5 u
22	116.8	1.8	576	5	Abv54556 Human pro
23	105.2	1.6	8056	7	Abz10246 Haematopo

24	101.4	1.5	8056	7	AbZ10246	Haematopo	
c	25	98.8	1.5	8056	7	AbZ10100	Haematopo
26	95	1.4	888	7	ACA56987	Human adi	
c	27	93.6	1.4	255	5	AaH82068	Rat diffe
28	86.6	1.3	8056	7	AbZ10100	Haematopo	
29	85.4	1.3	254	3	Aac27984	Human sec	
30	82	1.2	11805	6	AbL33748	Human imm	
31	76	1.1	251	3	AcO7640	Human sec	
32	75.6	1.1	7597	6	AbL33013	Human imm	
33	75.2	1.1	7657	4	Aa545477	Chemical	
34	75.2	1.1	7657	6	AbL34022	Human imm	
35	75.2	1.1	115218	7	AcA64845	Human HNR	
36	75	1.1	11812	4	AcA5502	Chemical	
37	75	1.1	11812	4	AcA46742	Tumour su	
38	75	1.1	11812	6	AbL34119	Human imm	
39	75	1.1	11812	6	AbK28432	DNA trans	
40	74	1.1	61020	4	AcA46788	Tumour su	
c	41	73.4	1.1	115218	7	AcA64845	Human HNR
42	73.2	1.1	5689	4	AcA5384	Chemical	
43	73.2	1.1	5689	4	AcA46426	Tumour su	
44	73.2	1.1	5689	6	AbK28426	DNA trans	
45	73	1.1	5930	6	AbL32517	Human imm	

ALIGNMENTS

RESULT 1

AAZ50069
ID AAZ50069 standard; cDNA; 6632 BP.

XX
AC AAZ50069;

DT 04-MAY-2000 (first entry)

XX DE Human Smad Anchor for Receptor Activation protein-2 encoding cDNA.

XX KW Smad Anchor for Receptor Activation protein; hSARA2; human;

XX KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;

XX KW BMP; activin; anti-inflammatory; cytostatic; antiarthritic; vulnery;

XX KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;

XX KW inflammatory response; tumour progression; cell proliferation; fibrosis;

XX KW fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX OS Homo sapiens.

XX PH Location/Qualifiers

XX FT CDS

XX FT /tag= a

XX FT /product= "Human SARA-2 protein"

XX FT /note= "Binds to receptor regulated Smad proteins"

XX FT /transl_except= (pos:336..338, aa:Gln)

XX FT /transl_except= (pos:822..824, aa:Thr)

XX RN WO200005360-A1.

XX PD C3-FES-2000.

XX PP 20-JUL-1999; 99WO-CA000656.

XX PR 20-JUL-1998; 98CA-02237701.

XX PR 10-DEC-1998; 98CA-02253647.

XX PA (HSCR-) HSC RES & DEV LP.

XX PI Wraha JL;

XX XX WPI; 2000-182691/16.

XX DR P-PSDB; AAY44750.

XX XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for

XX PT developing agents for treating e.g. wounds, arthritis, fibrosis,

XX PT inflammation, tumors, fibrogenesis or tissue damage.

XX Claim 8; Page 53-55; 93pp; English.

XX The present sequence is the cDNA encoding the human Smad Anchor for
 CC Receptor Activation protein-2 (hsARA2), a TGF-beta modulator. This has
 CC been identified in the EST clone, KIA0305 and is related to hSARA1. The
 CC hSARA proteins bind to receptor-regulated Smad proteins and ensures its
 CC appropriate localisation for activation by a type I receptor of a TGF-
 CC beta, activin or bone morphogenetic protein (BMP) signalling pathway.
 CC SARA proteins has anti-inflammatory, cytoskeletal, anti-arthritis and
 CC vulnary activity. The SARA protein is useful for developing agents for
 CC prevention or treatment of disorders involving TGF-beta superfamily
 CC member signalling pathway. Such agents can be used for wound healing,
 CC scarring, arthritis and fibrosis (e.g., liver and kidney), in modulating
 CC inflammatory and immune responses, tumour progression, cell proliferation
 CC or fibrogenesis and in tissue morphogenesis. It is also used for
 CC protecting, restoring and regenerating tissues after tissue damage

XX Sequence 6632 BP; 2197 A; 1123 C; 1253 G; 2059 T; 0 U; 0 Other;

Query Match 100.0%; Score 6632; DB 3; Length 6632;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACTCCCGCGGGGTAGCTTCTCACTCTCAGCGGAGCTGCTGTCGAGTTCCCAAAAG 60
 DB 1 ACTCCCGCGGGGTAGCTTCTCACTCTCAGCGGAGCTGCTGTCGAGTTCCCAAAAG 60

QY 61 CTCGCGAGGGGTAGGAGGTGATCTCATCAATTAAACAGCTGTGTGTCAGTTCCC 120
 DB 61 CTCGCGAGGGGTAGGAGGTGATCTCATCAATTAAACAGCTGTGTGTCAGTTCCC 120

QY 121 AAATCTTTATCTATCTCAGACTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 DB 121 AAATCTTTATCTATCTCAGACTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180

QY 181 TTGGATATCTCTCCAGGATGTTCTCAAGGATCAAGAAATTAATCTGTAATAGCTG 240
 DB 181 TTGGATATCTCTCCAGGATGTTCTCAAGGATCAAGAAATTAATCTGTAATAGCTG 240

QY 241 CAGGTAGATGACAGCTTATTTAAAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 241 CAGGTAGATGACAGCTTATTTAAAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 301 ATTTTGAACAGAACCCAGATGAACAGATTAATCTCGCAGATGTAACAAATGCATATG 360
 DB 301 ATTTTGAACAGAACCCAGATGAACAGATTAATCTCGCAGATGTAACAAATGCATATG 360

QY 361 CTAAACATGCTCAGTTTCTCAGAGTTGGCTTCTCTACAGGAACTTCATGTCCTCAA 420
 DB 361 CTAAACATGCTCAGTTTCTCAGAGTTGGCTTCTCTACAGGAACTTCATGTCCTCAA 420

QY 421 AAGACCAAGAGTGGCTTAATAGTTGCTCATCAGAAACAGCTATGGAACTAATGAGA 480
 DB 421 AAGACCAAGAGTGGCTTAATAGTTGCTCATCAGAAACAGCTATGGAACTAATGAGA 480

QY 481 GTTCCCTGAATGAAACAACTCAAGGACTTACTTCTATACAAATGAAACAAATGTA 540
 DB 481 GTTCCCTGAATGAAACAACTCAAGGACTTACTTCTATACAAATGAAACAAATGTA 540

QY 541 CAGGACTTGAATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 541 CAGGACTTGAATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

QY 601 ATATGGGACGATGATGATAAACCCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 601 ATATGGGACGATGATGATAAACCCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

QY 661 ATGCAACCAATAGTGAAGAGATTAATAAATTAATGCGCAGATGATTTAAGTCTAATG 720
 DB 661 ATGCAACCAATAGTGAAGAGATTAATAAATTAATGCGCAGATGATTTAAGTCTAATG 720

QY 721 CAGATTCTTGATTGGATTGATTATCTTCAAGTGTGAGATCTCCCTGTGTTCTTCAA 780

DB 721 CAGATTCTTGATTGGATTGATTATCTTCAAGTGTGAGATCTCCCTGTGTTCTTCAA 780

QY 781 CAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 781 CAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

QY 841 ATAGAGAAATCGAGAGATCAAGAAATCGGATATTAAGTAGATACACACTTTGAGATT 900
 DB 841 ATAGAGAAATCGAGAGATCAAGAAATCGGATATTAAGTAGATACACACTTTGAGATT 900

QY 901 CCTATATTTACAGTGGAAACAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960
 DB 901 CCTATATTTACAGTGGAAACAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAT 960

QY 961 CAATTCTGATTTTAAACATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 DB 961 CAATTCTGATTTTAAACATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1021 CCAAGACAGAGCTTACACACAGAGCCGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 DB 1021 CCAAGACAGAGCTTACACACAGAGCCGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

QY 1081 TAAAGAGGAGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 DB 1081 TAAAGAGGAGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140

QY 1141 AGACAAGTCTTGTGACCTGACCTTCCGAAATTAAGATGCTTAAATGATTTCAA 1200
 DB 1141 AGACAAGTCTTGTGACCTGACCTTCCGAAATTAAGATGCTTAAATGATTTCAA 1200

QY 1201 ATTCAAGAGATGAAATTTCAAAATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 DB 1201 ATTCAAGAGATGAAATTTCAAAATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

QY 1261 TAAACAATCTGCAACAGAGACTCAAAAGTTTACCTTAAGGATTAATGATTAATCC 1320
 DB 1261 TAAACAATCTGCAACAGAGACTCAAAAGTTTACCTTAAGGATTAATGATTAATCC 1320

QY 1321 AAGATTCCTTTCAGCTTTTACATGTTTCCAGTAAAGATGCTGCTGCTGCTGCTGCTG 1380
 DB 1321 AAGATTCCTTTCAGCTTTTACATGTTTCCAGTAAAGATGCTGCTGCTGCTGCTGCTG 1380

QY 1381 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 DB 1381 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1441 TACCTCAGCATGAACATGAAGATATATACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 DB 1441 TACCTCAGCATGAACATGAAGATATATACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

QY 1501 AGAACAGTGTGCTTCTAGTGGGGAAACCAATCAAGAGATGATCTTTGAAACAGGAAA 1560
 DB 1501 AGAACAGTGTGCTTCTAGTGGGGAAACCAATCAAGAGATGATCTTTGAAACAGGAAA 1560

QY 1561 AATGTAAGAGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 DB 1561 AATGTAAGAGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620

QY 1621 CTGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 DB 1621 CTGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680

QY 1681 TAGAATCTCAAGAGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 DB 1681 TAGAATCTCAAGAGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740

QY 1741 GTTTTATTAATGATTTTCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 DB 1741 GTTTTATTAATGATTTTCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1800

QY 1801 TTGATGAAGCGCAAAAGTGGCCCACTTAATTAAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 1860

Db 1801 TTGATGAAGGGCGAAAGATGGCCCACTAATTAGTGTGATGCTGAACCTTGATGCCCTTTCTGA 1860
Qy 1861 CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAATGTAAATGACTCTA 1920
Db 1861 CAGAACAGTATCTTCAGACCACTAAACATAAAGTCTTTTGAAGAAATGTAAATGACTCTA 1920
Qy 1921 AATCGCAAAATGAATCAGATAGATGAAGGCTTAGATGATGAAACATCAATAATATAT 1980
Db 1921 AATCGCAAAATGAATCAGATAGATGAAGGCTTAGATGATGAAACATCAATAATATAT 1980
Qy 1981 ATTCTAATGCAAGCAGGAGCTTATTTGGGAAAAGTCATGCTATTAATAATATTTGTGAAA 2040
Db 1981 ATTCTAATGCAAGCAGGAGCTTATTTGGGAAAAGTCATGCTATTAATAATATTTGTGAAA 2040
Qy 2041 CAGTGTATTAACAAATACATACATAGAAATGGCTTTCTTTAGCAGAAAAGACACTATTTC 2100
Db 2041 CAGTGTATTAACAAATACATACATAGAAATGGCTTTCTTTAGCAGAAAAGACACTATTTC 2100
Qy 2101 CAGTTCAAACAGGGTTACCTACCAGTAAGTCTGAGATTACAAATCAATATATCAGTCTCTG 2160
Db 2101 CAGTTCAAACAGGGTTACCTACCAGTAAGTCTGAGATTACAAATCAATATATCAGTCTCTG 2160
Qy 2161 ATATTAAACAGTCAATCTGTGTGGAGGGCCAGACCTTAAGCAATTTGTTAGCCTTCCATCAA 2220
Db 2161 ATATTAAACAGTCAATCTGTGTGGAGGGCCAGACCTTAAGCAATTTGTTAGCCTTCCATCAA 2220
Qy 2221 GAACAGGAGTTCAAAGSACTGTAATAAGCCAGATGTTCCAGATACAAATAGAAAGTGAAC 2280
Db 2221 GAACAGGAGTTCAAAGSACTGTAATAAGCCAGATGTTCCAGATACAAATAGAAAGTGAAC 2280
Qy 2281 CCAGCACAGCAGATACCGTTGTTTCCAAATCACTGTGCTATAGATTCTACAGCTGATCCAC 2340
Db 2281 CCAGCACAGCAGATACCGTTGTTTCCAAATCACTGTGCTATAGATTCTACAGCTGATCCAC 2340
Qy 2341 AGGTAGCTTCAACTCTAATTACATTCATATAGAAAGTAACTCTGAGGTGGATCTAGTT 2400
Db 2341 AGGTAGCTTCAACTCTAATTACATTCATATAGAAAGTAACTCTGAGGTGGATCTAGTT 2400
Qy 2401 TCGTAACTGCAAAATGAAGATTCTGTACTGAAACACCTTGCAAGAGGCTTGGTTTGG 2460
Db 2401 TCGTAACTGCAAAATGAAGATTCTGTACTGAAACACCTTGCAAGAGGCTTGGTTTGG 2460
Qy 2461 GCCAGAACAGCTTACTTGGTTCCTGATTCAGAGCTCCAACTGTATGAATGCCAAG 2520
Db 2461 GCCAGAACAGCTTACTTGGTTCCTGATTCAGAGCTCCAACTGTATGAATGCCAAG 2520
Qy 2521 TCAAAATTTACTTTTACCAACGGCGACACCACTCCGAGCATGTGGAAAAGTATTTTGG 2580
Db 2521 TCAAAATTTACTTTTACCAACGGCGACACCACTCCGAGCATGTGGAAAAGTATTTTGG 2580
Qy 2581 GTGCTCTGTGTAATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATAGTG 2640
Db 2581 GTGCTCTGTGTAATAGGAAGTGTAAACTGCAATATCTAGAAAAGGAAGCAAGATAGTG 2640
Qy 2641 TAGTCTGCTATGAACCTATAGTAAAGCTCAGGCATTTGAAAGGATCATGATGCCAACTG 2700
Db 2641 TAGTCTGCTATGAACCTATAGTAAAGCTCAGGCATTTGAAAGGATCATGATGCCAACTG 2700
Qy 2701 GTTCTAATCTTAAAGTCTAATCATTTCTGATGAATGTACTGTCCAGCCTCCTCAGGAGA 2760
Db 2701 GTTCTAATCTTAAAGTCTAATCATTTCTGATGAATGTACTGTCCAGCCTCCTCAGGAGA 2760
Qy 2761 ACCAAAACATCAGTATACCTTCACAGCACTTTGCCAGTCTCAGCACTTAAACAAACAG 2820
Db 2761 ACCAAAACATCAGTATACCTTCACAGCACTTTGCCAGTCTCAGCACTTAAACAAACAG 2820
Qy 2821 GTGTTGAAGGACTATGTTTCCAAAAGAACAGAGAGATGTTTGCAGATGGTATATTCG 2880
Db 2821 GTGTTGAAGGACTATGTTTCCAAAAGAACAGAGAGATGTTTGCAGATGGTATATTCG 2880
Qy 2881 CCAATGGTGAAGTTCAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTTCTGAAG 2940
Db 2881 CCAATGGTGAAGTTCAGATACAAACAAATTTATCATCTGGAAGTAAAGATGTTTCTGAAG 2940

Qy 2941 ACTTTAGTCTCTCTCACCTGATGTGCCTATGACAGTAAACACAGTGTGATCAITTCCTCAT 3000
Db 2941 ACTTTAGTCTCTCTCACCTGATGTGCCTATGACAGTAAACACAGTGTGATCAITTCCTCAT 3000
Qy 3001 CTACTACAGTGGAAAAGCCAAACAATGAGACAGGAGATATTACAGAAATGAGATAATTC 3060
Db 3001 CTACTACAGTGGAAAAGCCAAACAATGAGACAGGAGATATTACAGAAATGAGATAATTC 3060
Qy 3061 AGAGTCCCTATTCTCAGGTTCCATCAGTGGAAAATTTCTCTATGAACACAGAAATGAGG 3120
Db 3061 AGAGTCCCTATTCTCAGGTTCCATCAGTGGAAAATTTCTCTATGAACACAGAAATGAGG 3120
Qy 3121 GGTTCACCTACTCTCGTTCAITTTACACTAGATGATGTGTTTTGACAGAACTGAGAAC 3180
Db 3121 GGTTCACCTACTCTCGTTCAITTTACACTAGATGATGTGTTTTGACAGAACTGAGAAC 3180
Qy 3181 CATCTAGTCTCTACTGCTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATT 3240
Db 3181 CATCTAGTCTCTACTGCTCTTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTCAGATT 3240
Qy 3241 ATAGGTTACTGTGCTATTAACAAGTATGCTGCAATAAGATTAGTCTTCTACCTAATG 3300
Db 3241 ATAGGTTACTGTGCTATTAACAAGTATGCTGCAATAAGATTAGTCTTCTACCTAATG 3300
Qy 3301 ATGAGACAGTTTGGCCCACTCTCTGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG 3360
Db 3301 ATGAGACAGTTTGGCCCACTCTCTGTTGCAATCTGGAGAAAAGGATCAGTGCCTGTAG 3360
Qy 3361 TAGAAGACATCCATCTCATGAGCAGATCATTTCTGTTGAAAGTGAAGGCTTTCATC 3420
Db 3361 TAGAAGACATCCATCTCATGAGCAGATCATTTCTGTTGAAAGTGAAGGCTTTCATC 3420
Qy 3421 CTGTTACATTTGCTCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTATTCCT 3480
Db 3421 CTGTTACATTTGCTCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTATTCCT 3480
Qy 3481 CAGACAAATATTGGTACTTTTCAACCAATGATGCTGCTGGTGGGACAGGCAAAATTA 3540
Db 3481 CAGACAAATATTGGTACTTTTCAACCAATGATGCTGCTGGTGGGACAGGCAAAATTA 3540
Qy 3541 TTATTTCTATTGTTGTTGTTGCCAAATGAAGTACTATTCTTAAAGACATCTTCAGACTAT 3600
Db 3541 TTATTTCTATTGTTGTTGTTGCCAAATGAAGTACTATTCTTAAAGACATCTTCAGACTAT 3600
Qy 3601 TTATCACCATAATTAAGGATGCTTAAAGGAAAATACATAGAAAATTGGACAATATTA 3660
Db 3601 TTATCACCATAATTAAGGATGCTTAAAGGAAAATACATAGAAAATTGGACAATATTA 3660
Qy 3661 CCTTTACTGAGAGTTTCTCAGTAGCAAGATCAGGAGGATTCCTGTTTATTACACCTA 3720
Db 3661 CCTTTACTGAGAGTTTCTCAGTAGCAAGATCAGGAGGATTCCTGTTTATTACACCTA 3720
Qy 3721 CTTTTCAGAAAACCTTGATGATCTCTCATTAACCAAGTAACTCTTTCTTTGTGAAATTCCTA 3780
Db 3721 CTTTTCAGAAAACCTTGATGATCTCTCATTAACCAAGTAACTCTTTCTTTGTGAAATTCCTA 3780
Qy 3781 TCCAGAGCTTGAGATTCCCTGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGG 3840
Db 3781 TCCAGAGCTTGAGATTCCCTGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGG 3840
Qy 3841 GTGCAGAAATATAAGCATATCCTGCTCTTAAACAGCATCAGAGCCGCAAAACCTCTTT 3900
Db 3841 GTGCAGAAATATAAGCATATCCTGCTCTTAAACAGCATCAGAGCCGCAAAACCTCTTT 3900
Qy 3901 TTGGAGAAATAGACACACTATTATGAACCTTCTGTTGACCTTGAAATTTACCAAGTATA 3960
Db 3901 TTGGAGAAATAGACACACTATTATGAACCTTCTGTTGACCTTGAAATTTACCAAGTATA 3960
Qy 3961 CCTTGCATATATAGATCAACTGTTGATTCAATGAAATGGAAAAGCTGCATAAAAA 4020
Db 3961 CCTTGCATATATAGATCAACTGTTGATTCAATGAAATGGAAAAGCTGCATAAAAA 4020

QY 605 GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGACATGGGTAACTTAGTTCATGC 664
Db 468 GGGACGATGTAGTAAACCTATCTGTGATCTGATTAAGTGACATGGGTAACTTAGTTCATGC 527
QY 665 AACCAATAGTGAAGAAGATATTAATAAATTAATTCAGATGATTTAAGTCTAATGCAGA 724
Db 528 AACCAATAGTGAAGAAGATATTAATAAATTAATTCAGATGATTTAAGTCTAATGCAGA 587
QY 725 TTCCTTGATTTGATTTGATTTAATCTTCAGTGTGATCTACCTCCCTGTGTTCTTCAACAGA 784
Db 588 TTCCTTGATTTGATTTGATTTAATCTTCAGTGTGATCTACCTCCCTGTGTTCTTCAACAGA 647
QY 795 CCATGATAGTACATCTGTCAGAGAACACAGAAATGATATCAGTTCGAAATTTACAAAATAG 844
Db 648 CCATGATAGTACATCTGTCAGAGAACACAGAAATGATATCAGTTCGAAATTTACAAAATAG 707
QY 845 AGAAATCGGAGGAATCAAAAGATTTGGGTATATAAAGTAGATACAAACATTTCCAGATTCCTA 904
Db 708 AGAAATCGGAGGAATCAAAAGATTTGGGTATATAAAGTAGATACAAACATTTCCAGATTCCTA 767
QY 905 TAAATACAGTGAACAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGATCAAT 964
Db 768 TAAATACAGTGAACAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGATCAAT 827
QY 965 TGTTCATTTTAAATCTCATCTGCTTTGACTCGAAGATTTCCAAATGTTTCATGCAA 1024
Db 828 TGTTCATTTTAAATCTCATCTGCTTTGACTCGAAGATTTCCAAATGTTTCATGCAA 887
QY 1025 AGACAAGTACAAACAAGAGCCAGCCATGTGGATTAATCTAAAGATGTTGGCTTAGTAAA 1084
Db 888 AGACAAGTACAAACAAGAGCCAGCCATGTGGATTAATCTAAAGATGTTGGCTTAGTAAA 947
QY 1085 AGAGAAGTGTAGTGGCAGATCAATCTGCCGAGATGTTTAAAGAGAGGCGCAAGAC 1144
Db 948 AGAGAAGTGTAGTGGCAGATCAATCTGCCGAGATGTTTAAAGAGAGGCGCAAGAC 1007
QY 1145 AAGTCTTTGACCTGCAGCCTTCGAAAGATTAAGATTTATGCTTAAATGATCAAAATTC 1204
Db 1008 AAGTCTTTGACCTGCAGCCTTCGAAAGATTAAGATTTATGCTTAAATGATCAAAATTC 1067
QY 1205 AAGATGAAATTTCAATTTACCTGACTTTTCTTTTCAAGAGATGAGTCTGTTATAAA 1264
Db 1068 AAGATGAAATTTCAATTTACCTGACTTTTCTTTTCAAGAGATGAGTCTGTTATAAA 1127
QY 1265 ACAATCTGCACAAGAGACTCAAAAGATTTAGACCTTAAGGATTAATGATGTAATCCAAGA 1324
Db 1128 ACAATCTGCACAAGAGACTCAAAAGATTTAGACCTTAAGGATTAATGATGTAATCCAAGA 1187
QY 1325 TTCCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTGCGCTCATGTCCTGCTCTCC 1384
Db 1188 TTCCTCTTCAGCTTTACATGTTTCCAGTAAAGATGTGCGCTCATGTCCTGCTCTCC 1247
QY 1385 TCGCTCTGGTCTATGTGTGATCAATTAATGAAAGTAAAGACAGGGGTGATTTTATACC 1444
Db 1248 TCGCTCTGGTCTATGTGTGATCAATTAATGAAAGTAAAGACAGGGGTGATTTTATACC 1307
QY 1445 TCAGCATGAACATAAAGATATATACAGATGCGAGTACTATACATGAAGAAATACAGAA 1504
Db 1308 TCAGCATGAACATAAAGATATATACAGATGCGAGTACTATACATGAAGAAATACAGAA 1367
QY 1505 CAGTCTGTTCTAGTGGGAGCAATTCAGAGAGATGATCTTTTGAACAGGAAATG 1564
Db 1368 CAGTCTGTTCTAGTGGGAGCAATTCAGAGAGATGATCTTTTGAACAGGAAATG 1427
QY 1565 TAAAGCAGTACTCTTCAGTCAATTAATTAAGGGATGAAGACAGAAAGATAGATCTCGA 1624
Db 1428 TAAAGCAGTACTCTTCAGTCAATTAATTAAGGGATGAAGACAGAAAGATAGATCTCGA 1487
QY 1625 CCAGACAGTAAACAGAGTGTGCTTTGGATGGTGCACACAGTCTTACAGTTGTAGA 1684
Db 1488 CCAGACAGTAAACAGAGTGTGCTTTGGATGGTGCACACAGTCTTACAGTTGTAGA 1547
QY 1685 ATCTCAAGAGGGGCTTTCTGGCACTCATGTCCAGAGTCTTCTGATTTGTAAGGTTT 1744

Db 1548 ATCTCAAGAGGGGCTTTCTGGCACTCATGTCCAGAGTCTTCTGATTTGTAAGGTTT 1607
QY 1745 TATTAATACCTTTTCAAGCAATGATGATGGGCAAGACTTAGATTAATTAATGA 1804
Db 1608 TATTAATACCTTTTCAAGCAATGATGATGGGCAAGACTTAGATTAATTAATGA 1667
QY 1805 TGAAGCGCAAAAGTGGCCCACTAATTAGTATGATGCTCAACTTGTATGCTTTCGACAGA 1864
Db 1668 TGAAGCGCAAAAGTGGCCCACTAATTAGTATGATGCTCAACTTGTATGCTTTCGACAGA 1727
QY 1865 ACAGTATCTTCAGACCACTAATTAAGTCTTTTGAAGAAATGTAATGATCTTAATC 1924
Db 1728 ACAGTATCTTCAGACCACTAATTAAGTCTTTTGAAGAAATGTAATGATCTTAATC 1787
QY 1925 GCAATGATTCAGATGATGATGAAAGGCTTAGATGATGAAACATCAATATATATATTT 1984
Db 1788 GCAATGATTCAGATGATGATGAAAGGCTTAGATGATGAAACATCAATATATATATTT 1847
QY 1985 CAATGCAAGCAGGAGCTTATTTGGGAAAAGTCAATGATTAATTAATTTGTGAAACAGT 2044
Db 1848 CAATGCAAGCAGGAGCTTATTTGGGAAAAGTCAATGATTAATTAATTTGTGAAACAGT 1907
QY 2045 TGATAAACAAAATACAAATAGAAAATGGGCTTTCTTTGGAGAAAAGCACTATTTCCAGT 2104
Db 1908 TGATAAACAAAATACAAATAGAAAATGGGCTTTCTTTGGAGAAAAGCACTATTTCCAGT 1967
QY 2105 TCAACAAAGGTTTACCTTACCAGTAAGTCTGAGATTAACAATCAATTTATCAGTCTCTGATAT 2164
Db 1968 TCAACAAAGGTTTACCTTACCAGTAAGTCTGAGATTAACAATCAATTTATCAGTCTCTGATAT 2027
QY 2165 TAAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2224
Db 2028 TAAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2087
QY 2225 AAGGAGTTCAAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAAAAGTGAACCCAG 2284
Db 2088 AAGGAGTTCAAAAGGACCTGAATAAGCCAGATGTTCCAGATACAAATAGAAAAGTGAACCCAG 2147
QY 2285 CACAGCAGATACCGTTTCCAAATCACTTGTCTATAGATTTCTACAGCTGATCCACAGGT 2344
Db 2148 CACAGCAGATACCGTTTCCAAATCACTTGTCTATAGATTTCTACAGCTGATCCACAGGT 2207
QY 2345 TAGCTTCAATCTCAATTAATGATATAGAAAGTAACTGAGGCTGATCTAGTTTCGT 2404
Db 2208 TAGCTTCAATCTCAATTAATGATATAGAAAGTAACTGAGGCTGATCTAGTTTCGT 2267
QY 2405 AACTGCAAAATGAAGATTTCTGTACCTGAAAACACTTGCAAAAGAGCTTGGTTTGGGCCA 2464
Db 2268 AACTGCAAAATGAAGATTTCTGTACCTGAAAACACTTGCAAAAGAGCTTGGTTTGGGCCA 2327
QY 2465 GAAACAGCCTACTTGGGTTCTCGATTCAGAACCTCCAAACTGTATGAACCTGCAAGTCAA 2524
Db 2328 GAAACAGCCTACTTGGGTTCTCGATTCAGAACCTCCAAACTGTATGAACCTGCAAGTCAA 2387
QY 2525 ATTTACTTTTACCAACGGCGACACCATTTGCCAGCATGTGGGAAAGTATTTTGTGGTGT 2584
Db 2388 ATTTACTTTTACCAACGGCGACACCATTTGCCAGCATGTGGGAAAGTATTTTGTGGTGT 2447
QY 2585 CTGTTCTAATAGAAAGTGTAAATGCAATATCTAGAAAAGGAAAGCAAGATGTGTAGT 2644
Db 2448 CTGTTCTAATAGAAAGTGTAAATGCAATATCTAGAAAAGGAAAGCAAGATGTGTAGT 2507
QY 2645 CTGCTATGAACTATTTAGTAAAGCTCAGGCAATTTGAAAGGATGATGAGTCCAACTGGTTC 2704
Db 2508 CTGCTATGAACTATTTAGTAAAGCTCAGGCAATTTGAAAGGATGATGAGTCCAACTGGTTC 2567
QY 2705 TAACTTAAAGTCTAATCTTCTGATGAATGTACTACTGTCCAGCCTCTCAGGAGAACCA 2764
Db 2568 TAACTTAAAGTCTAATCTTCTGATGAATGTACTACTGTCCAGCCTCTCAGGAGAACCA 2627
QY 2765 AACATCAGTATACCTTCCACGCACTTTGCGAGTCTCAGCAGTAAACCAACAGGTTGT 2824

QY 4985 AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAATCATTTAGCACAAATTTAAATATC 5044
DB |||||
QY 4848 AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAATCATTTAGCACAAATTTAAATATC 4907
DB |||||
QY 5045 TAAAAATTAAGAGATCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAAAGA 5104
DB |||||
QY 4908 TAAAAATTAAGAGATCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAAAGA 4962
DB |||||
QY 5105 CAAGATTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTTATTAATGTCAGCCAAAG 5164
DB |||||
QY 4963 CAAGATTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTTATTAATGTCAGCCAAAG 5022
DB |||||
QY 5165 TATCTGCACTTAGGTATACCTCTTTTATGCGCAATTAATGATTTTAAATGAGGCTCTTTTCAG 5224
DB |||||
QY 5023 TATCTGCACTTAGGTATACCTCTTTTATGCGCAATTAATGATTTTAAATGAGGCTCTTTTCAG 5082
DB |||||
QY 5225 ATGTAACCTTTATGAGGAATATCTGCTTTGTGTATATGCGAGTTAGATCTGGTTTCT 5284
DB |||||
QY 5083 ATGTAACCTTTATGAGGAATATCTGCTTTGTGTATATGCGAGTTAGATCTGGTTTCT 5142
DB |||||
QY 5285 AAGCTCTGCAATTTGATTTTCACTGCGACAAACCAAGTTTTCAGGCTTTAGACTTATA 5344
DB |||||
QY 5143 AAGCTCTGCAATTTGATTTTCACTGCGACAAACCAAGTTTTCAGGCTTTAGACTTATA 5202
DB |||||
QY 5345 ATCTTTGAATAAACTGATAACTTATTTGTATATTTGAGTGGAGACCTACCTCCATAA 5404
DB |||||
QY 5203 ATCTTTGAATAAACTGATAACTTATTTGTATATTTGAGTGGAGACCTACCTCCATAA 5262
DB |||||
QY 5405 TTAGATAAACTCTTTTGGATTAATCAAGATTTTGGCTTTTCTTCTCAATATTTATA 5464
DB |||||
QY 5263 TTAGATAAACTCTTTTGGATTAATCAAGATTTTGGCTTTTCTTCTCAATATTTATA 5322
DB |||||
QY 5465 CATATGTATGTATATATATCCACATATATAGTTTCCCTGATTAATGATATTTAAAT 5524
DB |||||
QY 5323 CATATGTATGTATATATATCCACATATATAGTTTCCCTGATTAATGATATTTAAAT 5382
DB |||||
QY 5525 AA 5526
DB ||
QY 5383 AA 5384
DB ||

RESULT 3
AAC98041
ID AAC98041 standard; cDNA; 1572 BP.
XX AC AAC98041;
XX AC AAC98041;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:51.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.

OS Homo sapiens.

XX WO200055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005883.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX

DR WPI: 2000-587534/55.
DR P-PSDB; AAB53284.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer.
XX
PS Claim 1; Page 504-505; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins may
CC also be used to prevent diseases such as neural disorders, immune system
CC disorders, muscular disorders, reproductive disorders, gastrointestinal
CC disorders, wounds, renal disorders, infectious diseases, and
CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
CC sequences used in the exemplification of the present invention
XX

SQ Sequence 1572 BP; 537 A; 233 C; 231 G; 562 T; 0 U; 9 Other;

Query Match 22.8%; Score 1514; DB 3; Length 1572;
Best Local Similarity 99.3%; Pred. No. 5.9e-287;
Matches 1522; Conservative 7; Mismatches 3; Indels 1; Gaps 1;

QY 4865 TTAGTGAAGAATGTGCGCATATTCATATATGCAACCTAATTTGTTAAACCTAACCTCCAGC 4924
DB 1 TTAGTGAAGAATGTGCGCATATTCATATATGCAACCTAATTTGTTAAACCTAACCTCCAGC 60
QY 4925 ACTAAGCTGGAATGCCACAACTAAAAGTATATAATGTCTGATTTTGAACACAT 4984
DB 61 ACTAAGCTGGAATGCCACAACTAAAAGTATATAATGTCTGATTTTGAACACAT 120
QY 4985 AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAATCATTTAGCACAAATTTAAATATC 5044
DB 121 AAGCTTTGCTCTTTAGGAGGAGTATCTTTTCAATCATTTAGCACAAATTTAAATATC 180
QY 5045 TAAAAATTTAAGAGATCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAAAGA 5104
DB 181 TAAAAATTTAAGAGATCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAAAGA 240
QY 5105 CAAGATTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTTATTAATGTCAGCCAAAG 5164
DB 241 CAAGATTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTTATTAATGTCAGCCAAAG 300
QY 5165 TATCTGCACTTAGGTATACCTCTTTTATGCGCAATTAATGATTTTAAATGAGGCTCTTTTCAG 5224
DB 301 TATCTGCACTTAGGTATACCTCTTTTATGCGCAATTAATGATTTTAAATGAGGCTCTTTTCAG 360
QY 5225 ATGTAACCTTTATGAGGAATATCTGCTTTGTGTATATGCGAGTTAGATCTGGTTTCT 5284
DB 361 ATGTAACCTTTATGAGGAATATCTGCTTTGTGTATATGCGAGTTAGATCTGGTTTCT 420
QY 5285 AAGCTCTGCAATTTGATTTTCACTGCGACAAACCAAGTTTTCAGGCTTTAGACTTATA 5344
DB 421 AAGCTCTGCAATTTGATTTTCACTGCGACAAACCAAGTTTTCAGGCTTTAGACTTATA 480
QY 5345 ATCTTTGAATAAACTGATAACTTATTTGTATATTTGAGTGGAGACCTACCTCCATAA 5404
DB 481 ATCTTTGAATAAACTGATAACTTATTTGTATATTTGAGTGGAGACCTACCTCCATAA 540
QY 5405 TTAGATAAACTCTTTTGGATTAATCAAGATTTTGGCTTTTCTTCTCAATATTTATA 5464
DB 541 TTAGATAAACTCTTTTGGATTAATCAAGATTTTGGCTTTTCTTCTCAATATTTATA 600
QY 5465 CATATGTATGTATATATATCCACATATATAGTTTCCCTGATTAATGATATTTAAAT 5524
DB |||||

[illegible]

RESULT 4	
ABL8982	
ID	ABL89832 standard; cdna; 969 BP.
XX	
XX	
XX	ABL89832;
XX	
XX	
DT	24-MAY-2002 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 394.

KW	Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW	vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200190304-A2.
XX	
PD	29-NOV-2001.
XX	
PF	18-MAY-2001; 2001WO-US016450.
XX	
PR	19-MAY-2000; 2000US-020551SP.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Birse CE, Rosen CA;
XX	
DR	WPI; 2002-122018/16.
DR	P-PSDB; AB589423.
XX	
PT	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT	prevention of neural, immune system, muscular, reproductive,
PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT	disorders.
XX	
PS	Claim 4; SEQ ID NO 394; 2081pp + Sequence Listing; English.
XX	
CC	The invention relates to novel genes (ABL94949-ABL90853) and proteins
CC	(ABL99040-ABB90444) useful for preventing, treating or ameliorating
CC	medical conditions e.g. by protein or gene therapy. The genes are
CC	isolated from a range of human tissues disclosed in the specification.
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC	the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC	ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC	breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC	cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC	; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC	infectious diseases such as viral, bacterial, fungal and parasitic
CC	infections. Note: The sequence data for this patent did not form part of
CC	the printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SO	Sequence 969 BP: 294 A; 180 C; 177 G; 313 T; 0 U; 5 Other:

	Query Match	12.8%	Score 850,	DB 6,	Length 969,
	Best Local Similarity	97.1%	Pred. No. 7.6e-157,		
	Matches 915,	Conservative 1,	Mismatches 20,	Indels 6,	Gaps 5,
Qy	3093	AAATTGCTATGAACACAGGAAATGAGGGTTCACCTACTTCTGGTTCATTACACATAGAT	3152		
Db	1	AAATTGCTATGAACACAGGAAATGAGGGTTCACCTACTTCTGGTTCATTACACATAGAT	60		
Qy	3153	GATGATGTTTTTCAGAACTGAAGAACCAATCTAGTCCCTACTGGTGCTCTAGTTACAGC	3212		
Db	61	GATGATGTTTTTCAGAAACTGAAGAACCAATCTAGTCCCTACTGGTGCTCTAGTTACAGC	120		
Qy	3213	AAATTTACCTATTCTCTAGTATTTTCAGATTATAGTGTTACTGTGTGATATTAACAAGTATGTC	3272		
Db	121	AAATTTACCTATTCTCTAGTATTTTCAGATTATAGTGTTACTGTGTGATATTAACAAGTATGTC	180		
Qy	3273	TGCAATAAGATTAGTCTTCTACTCAATGATGAGGACAGTTTGCCCCCATCTCTGGTTGCA	3332		
Db	181	TGCAATAAGATTAGTCTTCTACTCAATGATGAGGACAGTTTGCCCCCATCTCTGGTTGCA	240		
Qy	3333	TCTGGAGAAAGGGATCAGTGCCCTGTGTAGTAGAAGAACATCCCATCTCATGACGAGATCAAT	3392		
Db	241	TCTGGAGAAAGGGATCAGTGCCCTGTGTAGTAGAAGAACATCCCATCTCATGACGAGATCAAT	300		

1213 AAAATTTCAAATACCTGACTTTTCCTTTCCAGGAATAAGACTGTATATAAAACAATCTG 1272
Cc
374 AAAATTTCAAATACCTGACTTTTCCTTTCCAGGAATAAGACTGTATATAAAACAATCTG 433
Cc
1273 CACAGAAGACTCAAAAGTTTACACCTTAAGGATATGATGTAATCCCAAGATTCCTCTT 1332
Cc
434 CACAAGAAGACTCAAAAGTTTACACCTTAAGGATATGATGTAATCCCAAGATTCCTCTT 493
Cc
1333 CAGCTTTACATGTTTCCAGTAAGATGTCGGTCTCATTTGCTGCTTCCTCCGCTCTG 1392
Cc
494 CAGCTTTACATGTTTCCAGTAAGATGTCGGTCTCATTTGCTGCTTCCTCCGCTCTG 553
Cc
1393 GGTCTATGTGTGATCAATTAATGAAGTAAGACACGGGGTGATTTTACCTCAGCATG 1452
Cc
554 GGTCTATGTGTGATCAATTAATGAAGTAAGACACGGGGTGATTTTACCTCAGCATG 613
Cc
1453 AACATTAAGATAATATACAGATGTCAGTGACTATACATGAAGAAATACAGACAGGTGTTG 1512
Cc
614 AACATAAAGATAATATACAGATGTCAGTGACTATACATGAAGAAATACAGACAGGTGTTG 673
Cc
1513 TTCTAGTGGGGAAC 1527
Cc
674 TTCTAGATGGGAAC 688
Cc

RESULT 6
ABQ57538
ID ABQ57538 standard; cDNA; 701 BP.
AC ABQ57538;
XX
XX
DT 02-AUG-2002 (first entry)
XX
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:1233.
XX
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
FN WO200229086-A2.
XX
XX
PD 11-APR-2002.
XX
XX
PF 02-OCT-2001; 2001WO-US030732.
XX
XX
PR 02-OCT-2000; 2000US-0237271P.
XX
XX
PA (FARB) BAYER CORP.
XX
XX
PI Burges C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiaglingam A, Lewis ME;
XX
XX
DR WPI; 2002-426115/45.
XX
XX

New isolated nucleic acid that is differentially expressed in cancer tissues useful for determining the presence of colon cancer in a cell or tissue type, and in antisense therapy.

Claim 1; Fig 1; 796pp; English.

ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be used in antisense therapy. An antibody immunoreactive with a polypeptide encoded by (I) is useful for detecting cancer in a patient sample, and for detecting the presence or absence of a polynucleotide encoded by a nucleic acid which hybridizes to (I) in a cell. A probe/primer derived from (I) can be used for determining the presence of a nucleic acid which hybridizes to (I), and for determining the phenotype of cells in a sample of cells from a patient. (I) is useful for determining the presence of colon cancer in a cell or tissue type, for determining the presence or

state of other type of cancer, in antisense therapy, to generate macroarrays on a solid surface, to identify a chromosome on which the corresponding gene resides, and in tissue profiling, forensics, genetic analysis, mapping and diagnostic applications. (I) can be used to raise antibodies, and to screen for peptide analogues and antagonists

Sequence 701 BP; 220 A; 134 C; 142 G; 197 T; 0 U; 8 Other;

Query Match 8.7%; Score 580.2; DB 6; Length 701;
Best Local Similarity 99.0%; Pred. No. 5.4e-104;
Matches 604; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

2736 ACTACTGTCAGCCCTCCTCAGAGAACCAACATCCAGTATACCTTCACCGACAACTTTG 2795
Cc
1 ACTACTGTCAGCCCTCCTCAGAGAACCAACATCCAGTATACCTTCACCGACAACTTTG 60
Cc
2796 CCAGTCTCAGCACTTAAACCAACAGGTGTGAAGACTATGTTCCAAAAGACAGAGA 2855
Cc
61 CCAGTCTCAGCACTTAAACCAACAGGTGTGAAGACTATGTTCCAAAAGACAGAGA 120
Cc
2856 GTATGTTTGCAGATGGTATATTTGCCAATGGTGAAGTTGCAGATACAAACAAATATCA 2915
Cc
121 GTATGTTTGCAGATGGTATATTTGCCAATGGTGAAGTTGCAGATACAAACAAATATCA 180
Cc
2916 TCTGGAAGTAAAGATGTTCTGAAGACTTTAGTCTCTCTCAGCTGATGCTCTATGACA 2975
Cc
181 TCTGGAAGTAAAGATGTTCTGAAGACTTTAGTCTCTCTCAGCTGATGCTCTATGACA 240
Cc
2976 GTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAAGCCAAACAAATGAGACAGGA 3035
Cc
241 GTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAAGCCAAACAAATGAGACAGGA 300
Cc
3036 GATATTACAGAAATGAGATTAATTCAGAGTCTTATTTCTCAGGTTCCATCAGTGGAAAAA 3095
Cc
301 GATATTACAGAAATGAGATTAATTCAGAGTCTTATTTCTCAGGTTCCATCAGTGGAAAAA 360
Cc
3096 TTGCTCTATCAACACAGGAAATCAGGGGTTACTACTCTGTTTCAATTTACACTAGATGAT 3155
Cc
361 TTGCTCTATCAACACAGGAAATCAGGGGTTACTACTCTGTTTCAATTTACACTAGATGAT 420
Cc
3156 GATGTTTTGAGAACTGAAGAACCATCTAGTCTCTACTGTTCTTGTCTTGTAGTTAAACAGCAAT 3215
Cc
421 GATGTTTTGAGAACTGAAGAACCATCTAGTCTCTACTGTTCTTGTCTTGTAGTTAAACAGCAAT 480
Cc
3216 TTACCTATTGCTAGTATTTCAGATTATAGTTTACTGTTGATATTAAAC- AAGTATGTTCTG 3274
Cc
481 TTACCTATTGCTAGTATTTCAGATTATAGTTTACTGTTGATATTAAACAGATTATGTTCTG 540
Cc
3275 CAATAAGATTAGTCTTCTACCTAATGATG-AGGACAGTTTCCCCCCTTCTGTTGAT 3333
Cc
541 CAATAAGATTAGTCTTCTACCTAATGATGAGGACAGTTTCCCCCCTTCTGTTGAT 600
Cc
3334 CTGGAGAAAA 3343
Cc
601 CTGGAGAAAA 610
Cc

RESULT 7
AAZ50068

ID AAZ50068 standard; cDNA; 4839 BP.

XX AAZ50068;

XX AAZ50068;

DT 04-MAY-2000 (first entry)

XX Human Smad Anchor for Receptor Activation protein-1 encoding cDNA.

Smad Anchor for Receptor Activation protein; hSAR1; human;
transforming growth factor-beta; TGF-beta; bone morphogenetic protein;
BMP; activin; anti-inflammatory; cytotactic; antiarthritic; vulnary;
TGF-beta modulator; wound healing; scarring; arthritis; immune response;
inflammatory response; tumour progression; cell proliferation; fibrosis;
fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX OS Homo sapiens.
 XX FH Key
 XX FT CDS
 XX FT Location/Qualifiers
 XX FT 439..4410
 XX FT /*tag= a
 XX FT /product= "Human SARA-1 protein"
 XX FT /note= "Binds to receptor regulated Smad proteins"
 XX FT /transl_except= (pos:4177..4179, aa:Thr)
 XX FT /transl_except= (pos:4180..4182, aa:Gln)
 XX FT /transl_except= (pos:4198..4200, aa:Gly)
 XX FT /transl_except= (pos:4279..4281, aa:His)
 XX FT /transl_except= (pos:4306..4308, aa:Phe)
 XX FT /transl_except= (pos:4315..4317, aa:Asp)
 XX FT /transl_except= (pos:4324..4326, aa:Leu)
 XX FT /transl_except= (pos:4329, aa:Met)
 XX PN WO200005360-A1.
 XX PD 03-FEB-2000.
 XX PF 20-JUL-1999; 99WO-CA000656.
 XX PR 20-JUL-1998; 98CA-02237701.
 XX PR 10-DEC-1998; 98CA-02253647.
 XX PA (HSCR-) HSC RES & DEV LP.
 XX PI Wraha JL;
 XX DR WPI; 2000-182691/16.
 XX DR P-PSDB; AAY44749.
 XX PT New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 XX PT developing agents for treating e.g. wounds, arthritis, fibrosis,
 XX PT inflammation, tumors, fibrogenesis or tissue damage.
 XX PS Claim 7; Page 50-51; 93pp; English.
 CC The present sequence is the cDNA encoding the human Smad Anchor for
 CC Receptor Activation protein-1 (hSAR1), a TGF-beta modulator. It is
 CC isolated from lambda ZAP human fetal brain cDNA library and has 62%
 CC identity to hSAR1. The hSAR1 proteins bind to receptor-regulated Smad
 CC proteins and ensures its appropriate localisation for activation by a
 CC Type I receptor of a TGF-beta, activin or bone morphogenetic protein
 CC (BMP) signalling pathway. SARA proteins has anti-inflammatory,
 CC cytoskeletal, anti-arthritis and vulnary activity. The SARA proteins are
 CC useful for developing agents for prevention or treatment of disorders
 CC involving TGF-beta superfamily member signalling pathway. Such agents can
 CC be used for wound healing, scarring, arthritis and fibrosis (e.g., liver
 CC and kidney), in modulating inflammatory and immune responses, tumour
 CC progression, cell proliferation or fibrogenesis and in tissue
 CC morphogenesis. It is also used for protecting, restoring and regenerating
 CC tissues after tissue damage
 XX SQ Sequence 4839 BP; 1460 A; 965 C; 1140 G; 1274 T; 0 U; 0 Other;
 Query Match 8.5%; Score 560.6; DB 3; Length 4839;
 Best Local Similarity 60.8%; Pred. No. 5.9e-100;
 Matches 990; Conservative 0; Mismatches 619; Indels 18; Gaps 4;
 3294 CCTAATGATGAGGACAGTTGCCCCCTCTGCTTCATCTGAGAGAAAGGATCATGTG 3353
 2824 CTTATCTCTGAAGATGGCCCTTCCCTCCATCTCTCCACTGGTGTAAAGG---AGC 2880
 3354 CCTGTAGTAGAAGAACATCCATCTCATGACAGATCATTTTGTCTTGTGAAGGTGAAGC 3413
 2881 TATGCTGTGGAAGAGAAACCATCATCAGATTTTCAGTAATGCAGCAGTTGGAGGATGGTGC 2940
 3414 TTTTATCTCTGTATCTTCTTAAGTCTATCTCTGATGTCATATTCATATTT 3473
 2941 CCTGACCCACTGTATTGTGTTTAAATGCAAAATTTGTTGTCATGGTTAAATGTGTAAT 3000

QY 3474 TATTCCTCAGACAAATATTGGTACTTTTCAACCAATGATTGCTGCTTGGACAGGCA 3533
 DB 3001 TATGTGAACAGGAAGTGTGTGTGTTTCAACCAAGGGAATGCATGAGTGGTCACTCT 3060
 QY 3534 GAAATTATTATTCTATTGTTATGTTGGCAATGAAGATCTATTCTTAAGGACATCTTC 3593
 DB 3061 GAGATAGTCAATCTTCTACAGTGTTCACCGATGAAAGTGTTCGCCAAGGATATCTTT 3120
 QY 3594 AGACTATTATTACCATATATATAGGATCTCTAAAGGAAATACATAGAAAATCTTGAC 3653
 DB 3121 AATCACTTTGTGAGCTTTATCGGATGCTCTGCGAGGAATGTGGTGAACATTTGGGA 3180
 QY 3654 AATATTACCTTTACTAGAGATTTTCTCAGTAGAAGATCAGGAGATTCCTCTTTATT 3713
 DB 3181 CATTCTCTTCTCAGTCAAAAGTTTCTTGGCAGTAAAGAACATGTTGGATTTCTTATATGTG 3240
 QY 3714 ACACCTACTTTTCAGAAAATTTGATGATCTCTATACCAAGTAATCTTTTCTTTGGGA 3773
 DB 3241 ACATCTACTACCACTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTTTGGG 3300
 QY 3774 ATTCTTATCCAGAACCTTTGAGATTCCTTGGGCAAGGTTTTTCTATGCGTTTAAATGTG 3833
 DB 3301 ATTCTTATCCAGAAATGGAACCTCTTGGGCTAAAGTATTTCTATCCGTCATGTGTG 3360
 QY 3834 AGATTGGGTGAGAAATATAAAGCATATCTCTCTCTTAACAAGCATCAGAGGCCGAAA 3893
 DB 3361 AGACTTGGAGCTGAATATCGACTTTATCCATGCCCACTTAITTCAGTGTCAAGATTCGGAAG 3420
 QY 3894 CCTCTTTTGGAGAAATAGGACACACTATTATGAATTTTCTTGTGCTTTGACCTTCGAAATATAC 3953
 DB 3421 CCATTGTTTGGAGACCGGGCATACCATCATGAATCTTCTTTCAGAGATTCAGAAATATAC 3480
 QY 3954 CAGTATACCTTTCGAPATAATATAGATCAACTGTGTTGATTCATATGAAATGGGAAAAGCTGC 4013
 DB 3481 CAGTATACCTTTCGAGTAGTTTCAAGTTTGGTGGTGTATGTAAGAGTTCGGAATATAGC 3540
 QY 4014 ATAAAAATACACGGAAGAAAGTACAGTGTATTAAGTAAAGTACTAAATTTCTTCAATGAG 4073
 DB 3541 ATCAAAATTCGAGCAACAGATACATGAGATGATGAAGCCATGAACAAGTCCATGAG 3600
 QY 4074 CATGTCAATAGCATTTGGAGCAAGTTTTCAGTACAGAAAGCAGATTTCTCATCTAGTCTGTATA 4133
 DB 3601 CATGTCTCGCAGGAGGTGCGCTGCTTCAATGAAAGGCAGACTCTCATCTTGTGTGTGTA 3660
 QY 4134 CAGA---ATGATGGAATTTTGAACACAGCCACAGTGCACCTGGCCATCTCTAGAAA 4190
 DB 3661 CAGAAATGATGTTGGAACACTATCAGACCAGGCTATCAGTATTTCATTCAGCCAGAAA 3720
 QY 4191 GTGACAGGTGCAAGTTTGTGTATTCAATGGAGCTCTAAAAACATCTTCAGGATTTCTT 4250
 DB 3721 GTGATGTGTGCAGTTTCTTGTGTTTCACTGGCGCTCTGAATCTCTTCTTGGATACCTT 3780
 QY 4251 GCTAAGTCCAGCATAGTTGAAGATGGCTTAAATGTTAAGTAAATTAATCTCCAGAGACCATGAAT 4310
 DB 3781 GCGAAGTCCAGTATTGTGGAAGATGTTTATGTTGTTCCAGATTACTGACAGAAATGGAT 3840
 QY 4311 GCGTTGCGGCTAGCTTTTACGAGAACAGAAAGACTTTTAAATTTACATGTTGGAAAGTTGAT 4370
 DB 3841 TCCTTTGAGCGAGGCACTCGGAGAGATGAAGACTTCACCATCACTGTGTGGAAAGCGGAC 3900
 QY 4371 GCAGTAGACCTGAGAGAAATACGTGGATATCTGTGGGTAGATGCTGAAGAAAAGGAAAC 4430
 DB 3901 GCGGAGGAACCCAGGAGCACATCCACATCCAGTGGGTGGATGATGACAGAAAGCTTAGC 3960
 QY 4431 AAAGAGTATACAGTTTCACTGGTGAATATCATTTACAAGGATTTTCCAAAGTGAAGAAATA 4490
 DB 3961 AAGGGTGTGTAAGTCTTATAGATGGGAAGTCCATGGAGACTATTAACAATGTGAAGATA 4020
 QY 4491 AAATCTGGAAGCAGATTTTGAACCCGATGAGAAGATTTGTAATATGTACCGAGGTGTTCTAC 4550
 DB 4021 TTCCATGGATCAGATATATAAGCAATTAAGAAAGTAATCAGATGACACAGAGGTGTTTTTC 4080
 QY 4551 TTTCTAA-----AGACCAGGATTTATCTATTATTTTATCAACTTCTTATCAGTTTGA 4601

Db	4081	CTAGAAACCGATGACCGACCAATTGGCTCAGTGAATCCTCGAGATCACATGATTGACT	4140
Qy	4502	AAAGAAATAGCCATGGCTTGTAGTGTGCGCTGTGCGCTCACCTCGAAAACTCTAAAAAGT	4561
Db	4141	GAGCATGTCGCAAAAGCTTTTGGCTTGGCTCTCTGTCTCCTCACCTGAAACTTCTGAAGAA	4200
Qy	4562	ANTGGGATGAATAAATTTGGACTCAGAGTTTCAATTGACACTGATATGGTTGAATTTCAG	4721
Db	4201	GATGGAATGACCAAACTGGGACTACGTGTGACACTTGACATCAGATCAGGTGGCTATCAA	4260
Qy	4722	GCAGGATCTGAAGGCCAACTTTCTGCCTCAGCATTATCTAAATGATCTTGATGTGCTCTG	4781
Db	4261	GCAGGAGCAATGSCCAGCCCCITCCCTCGCAGTACATGAATGATCTGGATAGCGCCTTG	4320
Qy	4782	ATACCTGTGATCCATGTGTGGAGCTTCCAATCTTAGT---TTACCATTAGAAATAGAAATTA	4838
Db	4321	GTGCGGTGTATCCATGTAGAGGGGCTGCCAGCTTAGTGAGGGCCCGGTGTGATCGGAATC	4380
Qy	4839	GTGTTTTTCATTATAGAAACATCTTTTTTAGTGAAGAATGTGCCATATTACATATTGCAA	4898
Db	4381	ATCTTTTATATCTGGAAACATCGTATAACAGAGAAGACTTCATTTTTTCTGTTCAG	4440
Qy	4899	CCCTAAAT	4905
Db	4441	ACTGTT	4447

RESULT 8
ADC64443
ID ADC64443 standard; DNA; 1977 BP.
XX
XX
AC ADC64443;
XX
XX
DT 18-DEC-2003 (first entry)
XX
XX
DE Human SARA_v5 coding sequence.
XX
XX Hepatotropic; vaccine; protein-protein interaction;
KW Transforming Growth Factor beta; TGF beta; hepatitis;
KW Selected Interacting Domain; SID; bait; human; SARA v5; gene; ds.

New complex between two interacting proteins, useful for screening PT molecules that inhibit transforming growth factor beta (TGF beta) or TGF beta super-family of cytokines pathway for diagnosing or treating TGF beta diseases or disorders.

XX PS Disclosure; SEQ ID NO 10; 148pp; English.

xx The present invention relates to protein-protein interactions and
 CC complexes involved in Transforming Growth Factor (TGF) beta disorders
 CC and/or diseases. The complex between two interacting proteins is useful
 CC for screening molecules that inhibit TGF beta for diagnosing or treating
 CC diseases or disorders involving TGF beta e.g., hepatitis. To illustrate

QY 3616 AGGATGCTCTAAAGGAAATACATAGAAAACTTGGCAATATTTACTTCTGAGAGTT 3675
 DB 121 AGGATGCTCTAAAGGAAATACATAGAAAACTTGGCAATATTTACTTCTGAGAGTT 180
 QY 3676 TTCTCAGTAGAAGATCAGGAGGATTCCTGTTTATTTACACCTTCTGAGAACTTG 3735
 DB 181 TTCTCAGTAGAAGATCAGGAGGATTCCTGTTTATTTACACCTTCTGAGAACTTG 240
 QY 3736 ATGATCTCTCATTAACCAAGTAATCTTTCTTTGGAATTTCTTATCCAGAGCTTCGAGA 3795
 DB 241 ATGATCTCTCATTAACCAAGTAATCTTTCTTTGGAATTTCTTATCCAGAGCTTCGAGA 300
 QY 3796 TTCCCTGGGCAAGGTTTTCTTATGCGTTTAAATGTTGAGATTCGGTGCAGAAATATAAG 3855
 DB 301 TTCCCTGGGCAAGGTTTTCTTATGCGTTTAAATGTTGAGATTCGGTGCAGAAATATAAG 360
 QY 3856 CATATCTCTCTCTATACAGGATCAGAGCGGCAAACTCTTTTGGAGAAATAGGAC 3915
 DB 361 CATATCTCTCTCTATACAGGATCAGAGCGGCAAACTCTTTTGGAGAAATAGGAC 420
 QY 3916 ACACATATTATGAACCTTACTTCTGACCTTCGAAATTTACAGTATACCTTTGCATA 3969
 DB 421 ACACATATTATGAACCTTACTTCTGACCTTCGAAATTTACAGTATACCTTTGCATA 474

RESULT 11

AAH81649/c
 ID AAH81649 standard; DNA; 455 BP.

AC AAH81649;

DT 21-SEP-2001 (first entry)

DE Human differential transcription-associated cDNA SEQ ID 158.

KW Differential transcription; human; rat; tumour cell; cytostatic;

KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

OS Homo sapiens.

PN WO200157058-A2.

PD 09-AUG-2001.

PF 31-JAN-2001; 2001WO-EP001003.

PR 31-JAN-2000; 2000DE-01004102.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

PI Grips M, Hellriegel M, Schmitz A, Sers C;

DR WPI; 2001-483415/52.

PT Nucleic acids differentially expressed between tumor and normal cells,

PT useful for diagnosis or therapy of tumors and for screening active

PT agents.

XX Disclosure; Page 372; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression
 CC between tumour and normal cells and which has cytostatic activity. (I)
 CC work as modulators of Ras activity by inducing expression of tumour
 CC suppressor genes. (I), and polypeptides encoded by them, are useful as
 CC targets for diagnosis or therapy and in screening to determine the
 CC effects of an active compound (potential pharmaceutical) on a cell line,
 CC particularly for diagnosis and treatment of tumors, especially by
 CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
 CC methods) or by modulating the amount and/or location of (I)-encoded
 CC polypeptides (by administration of the polypeptide or its activator,
 CC antibody (optionally as a conjugate) or inhibitor). The method allows
 CC identification of many Class II tumour suppressor genes (i.e. genes that

CC are not primary targets for tumour-initiating mutations). AAH81492-
 CC AAH82376 represent the human and rat derived nucleic acid fragments
 CC described in the method of the invention

XX Sequence 455 BP; 185 A; 61 C; 70 G; 139 T; 0 U; 0 Other;

Query Match 6.7%; Score 444; DB 5; Length 455;

Best Local Similarity 99.8%; Pred. No. 2.4e-77;

Matches 455; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 5337 GACTTATATTTCTTTGAATAAACTGATAACTTATTGTATTAATTTGGAGTGAGACCTTAC 5396

DB 455 GACTTATATTTCTTTGAATAAACTGATAACTTATTGTATTAATTTGGAGTGAGACCTTAC 396

QY 5397 CTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTCTTCTCA 5456

DB 395 CTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTCTTCTCA 336

QY 5457 AATTATACATATGATGATTTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 5516

DB 335 AATTATACATATGATGATTTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 276

QY 5517 ATTAATAAATTTGCGGGTCTTCAGGACTTTTGGCTTCTATATTTAAGTATATGTTTTT 5576

DB 275 ATTAATAAATTTGCGGGTCTTCAGGACTTTTGGCTTCTATATTTAAGTATATGTTTTT 216

QY 5577 ATAGCAAGAACATATCTGAAATGTTTTTATAAATCTTTAATATATATAGTATAGTAAAT 5636

DB 215 ATAGCAAGAACATATCTGAAATGTTTTTATAAATCTTTAATATATATAGTATAGTAAAT 156

QY 5637 TTTTGTATCACAAATGCAATATTTTTCCTCTCTTCCCTTCCAACTATACCACTGAT 5696

DB 155 TTTTGTATCACAAATGCAATATTTTTCCTCTCTTCCCTTCCAACTATACCACTGAT 97

QY 5697 TACCACCTTCTAAGAGTGAGTACGACGCGGCCAGATGACCTTTGAAGTAGTATGATAG 5756

DB 96 TACCACCTTCTAAGAGTGAGTACGACGCGGCCAGATGACCTTTGAAGTAGTATGATAG 37

QY 5757 CAATAAATGAAGCCTGAAACAGGTTTTTTTACTTCC 5792

DB 36 CAATAAATGAAGCCTGAAACAGGTTTTTTTACTTCC 1

RESULT 12

ACH20104
 ID ACH20104 standard; cDNA; 391 BP.

XX ACH20104;

XX 13-OCT-2003 (first entry)

DE Human adult lung cDNA #1107.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;

XX Genome mapping; biodiversity; genetic disorder.

XX Homo sapiens.

XX US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX DR WPI; 2003-615964/58.
XX PT
XX PT New polynucleotide sequences obtained from various cDNA libraries, useful
XX PT as hybridization probes, as oligomers for PCR, for chromosome and gene
XX PT mapping, in the recombinant production of protein, or in generating
XX PT antisense DNA or RNA.
XX PS Claim 1; SEQ ID NO 7316; 44pp; English.
XX CC
XX CC The invention relates to an isolated polynucleotide comprising any one of
XX CC 38043 cDNA sequences, appearing as ACHI2789-ACH50831, whose sequence was
XX CC determined by the technique of SBH (sequencing by hybridisation). Also
XX CC included is a purified polypeptide comprising a sequence corresponding to
XX CC a reading frame of the novel polynucleotide. The nucleic acid sequences
XX CC are useful in diagnostics as expressed sequence tags (EST) for
XX CC identifying expressed genes or for physical mapping of the human genome,
XX CC in forensics, in assessing biodiversity, or in identifying mutations
XX CC responsible for genetic disorders and other traits. The nucleotide
XX CC sequences are also useful as hybridisation probes, as oligomers for PCR,
XX CC for chromosome and gene mapping, in the recombinant production of
XX CC protein, or in generating antisense DNA or RNA. The purified polypeptide
XX CC is useful for generating antibodies specific for it. The present sequence
XX CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX CC for this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html?docID=20030073623
XX CC
XX CC Sequence 391 BP; 132 A; 71 C; 64 G; 124 T; 0 U; 0 Other;
XX CC
XX CC Query Match 5.9%; Score 389.4; DB 8; Length 391;
XX CC Best Local Similarity 99.7%; Pred. No. 1.2e-66;
XX CC Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX CC
Qy 4639 CTCACCTGAAACCTCTAAAGAGTAATGGATGATATTAATGGACTCAGAGTTTCCATTG 4698
Db 1 CTCACCTGAAACCTCTAAAGAGTAATGGATGATATTAATGGACTCAGAGTTTCCATTG 60
Qy 4699 ACACGTATATGGTTGCAATTTTCAGGCAGGATCTGAGGCCCAACTTCTGCCTCAGCATATC 4758
Db 61 ACACGTATATGGTTGCAATTTTCAGGCAGGATCTGAGGCCCAACTTCTGCCTCAGCATATC 120
Qy 4759 TAAATGATCTTGATGCTCTGATACCTGATCCATGGTGGACCTCCAACTCTAGTT 4818
Db 121 TAAATGATCTTGATGCTCTGATACCTGATCCATGGTGGACCTCCAACTCTAGTT 180
Qy 4819 TACCATTAGAATAAGATTTAGTGTCTTTTCATTATAGAACATCTTTTGTGAAAGATG 4878
Db 181 TACCATTAGAATAAGATTTAGTGTCTTTTCATTATAGAACATCTTTTGTGAAAGATG 240
Qy 4879 TGCCATATTACATATTGCAACTAATTTGTTTAAACTTAATCTCCAGCACTAAAGCTGAAT 4938
Db 241 TGCCATATTACATATTGCAACTAATTTGTTTAAACTTAATCTCCAGCACTAAAGCTGAAT 300
Qy 4939 GCCCAACACCTAAAGATTAATATGCTCATTTTGAACACATAAGCTTTGCTCTTT 4998
Db 301 GCCCAACACCTAAAGATTAATATGCTCATTTTGAACACATAAGCTTTGCTCTTT 360
Qy 4999 AGGCAGGAATGATCTTTTCAATCATTAGCA 5029
Db 361 AGGCAGGAATGATCTTTTCAATCATTAGCA 391
RESULT 13
AAH35041
ID AAH35041 standard; cDNA; 317 BP.
XX
XX AAH35041;
XX AC
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen encoding cDNA SEQ ID NO:2123.
XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR P-PSDB; AAG75636.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 3604-3605; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922
XX
SQ Sequence 317 BP; 97 A; 55 C; 66 G; 96 T; 0 U; 3 Other;
XX
XX Query Match 4.3%; Score 286; DB 4; Length 317;
XX Best Local Similarity 95.6%; Pred. No. 2.1e-46;
XX Matches 303; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
Qy 1609 GAAAGATAGATCTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGGACACCA 1668
Db 2 GAAAGATAGATCTGACACAGACAGTAATCAGAGCTGAGTCTTTGGATGGTGGACACCA 61
Qy 1669 GTTCTACAGTTGTAGAAATCTCAAGAGGGGCTTTCTGGCACTCATGTGCCAGAGTCTTCTG 1728
Db 62 GTTCTACAGTTGTAGAAATCTCAAGAGGGGCTTTCTGGCACTCATGTGCCAGAGTCTTCTG 121
Qy 1729 ATTGTTGTGAAGTTTTTATTATACTCTTTTCAAGCAATGATATGATGGCAGACTTAG 1788
Db 122 ATTGTTGTGAAGTTTTTATTATACTCTTTTCAAGCAATGATATGATGGCAGACTTAG 181
Qy 1789 ATTACTTTTAAATTTGATGAAGCGCAAAAGTGGCCCACTTAATTAGTGTGTAACCTTG 1848
Db 182 ATTACTTTTAAATTTGATGAAGCGCAAAAGTGGCCCACTTAATTAGTGTGTAACCTTG 240
Qy 1849 ATGCCCTTTCTGCAGAACAGATATCTTTCAGACCACTAACATAAGTCTTTTGAAGAAAATG 1908
Db 241 ATGCCCTTTCTGCAGAACAGATATCTTTCAGACCACTAACATAAGTCTTTTGAAGAAAATG 300
Qy 1909 TAAATGACTCTAAATCG 1925

Db 301 TAAATGACTCTAAATCG 317

RESULT 14
ABX48202
ID ABX48202 standard; cDNA; 393 BP.

XX
AC ABX48202;
DT 21-FEB-2003 (first entry)
XX

DE Bovine EST associated with lactation/muscle/fat deposition #13367.

XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX

OS Bos Taurus.
XX

PN US2002137139-A1.
XX

PD 26-SEP-2002.
XX

PF 24-SEP-2001; 2001US-00960352.
XX

PR 12-JAN-1999; 99US-0115707P.
PR 11-JAN-2000; 2000US-00480902.
XX

PA (BYATT/) BYATT J C.
PA (MATH/) MATHALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX

PI Byatt JC, Mathialagan N, Tao N, Warren WC;
WPI; 2003-110599/10.
XX

PT New nucleic acid associated with lactation, and muscle and fat
deposition, useful for genome mapping, gene identification and analysis,
cattle breeding, or for genetically improving cattle.
XX

PS Claim 2; SEQ ID NO 13367; 245pp; English.
XX

CC The invention relates to a purified nucleic acid molecule associated with
lactation or muscle and fat deposition (designated LMFD), derived from
cattle, and the LMFD nucleic acid can specifically hybridize to a second
nucleic acid molecule comprising any of 15112 nucleotide sequences,
appearing as ABX4836-ABX49947, or complements of them. Also included are
functions in the cell to cause termination of transcription and addition
of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
(2) determining a level or pattern of a molecule in a bovine cell or
tissue comprising: (a) incubating a marker nucleic acid (comprising any
of the 15112 nucleic acid sequences or its complement or fragment) with a
complementary nucleic acid molecule obtained from the bovine cell or
tissue, where hybridisation between the marker nucleic acid and the
complementary nucleic acid permits the detection of the molecule; and (b)
detecting the level or pattern of the complementary nucleic acid, where
the detection of the complementary nucleic acid is predictive of the
level or pattern of the molecule. The LMFD nucleic acid is used for
determining a level or pattern of a molecule in a bovine cell or tissue.
It is useful for genome mapping, gene identification and analysis, cattle
breeding, preparation of constructs for use in cattle gene expression, or
for genetically improving cattle. The present sequence is one of the
15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
present sequence was not shown in the specification but was obtained in
electronic format from the USPTO web site:
segdata.uspto.gov/sequence.html?docID=20020137139
XX

SQ Sequence 393 BP; 124 A; 53 C; 63 G; 153 T; 0 U; 0 Other;
Query Match 3.0%; Score 199.6; DB 7; Length 393;

Best Local Similarity 81.4%; Pred. No. 1.9e-29;
Matches 311; Conservative 0; Mismatches 49; Indels 22; Gaps 6;

QY 5111 TTTCTTTTAAGAAATTTATAGCATTTTACTGTGTTATTTAAATGCT-AAGCCAAAGTATCT 5169
DB 22 TGTCTTTATAGAAATTTTGTAGCATTTTACTGTGTTTAAATGCTTAAAGCCAAAGTATCT 81
QY 5170 GCATTTAGGTATACCTCTTTTATGCCAATAATGATTTTAAATGAAGCTCTTTTCAGATGTA 5229
DB 82 GCACT-----TTATGCCAAGAAGCATTTTATGTAAGGCTCTTTTCAGATGTA 129
QY 5230 ACCTATGAGGAATATCTGCTTTTGTGTATATGCCAGTTAGCACTTAACTGTTCTTAAAGT 5289
DB 130 ACCTATGAGGAATATCTGCTTTTGTGTATATGCCAGTTAAATTTACTGGTTTCTTAAAGT 189
QY 5290 CTG-TCAAATTTGATTTTCTAGTGGCACAACAAACAGTTTGTAGGCTTTAGACTT-----AT 5343
DB 190 CTGTTAAATATGATTTTCAATGGCAC--AGACCAGTTTTCAGATCTTACACTTATATGAT 247
QY 5344 AATCTTTGAATATAAATCTGA-TAACTTATTTGTATATGAAGTGGAGAGACCTACCTCCAT 5402
DB 248 AATCTTTGAATATAAATCTGA-TAACTTATTTGTATATGAAGTGGAGAGACCTACCTCCAT 307
QY 5403 AATTAGATAAATCTCTTTTGGATTTATTAATCAGAAATTTTGCCCTTTTCTTCAAAATAT 5462
DB 308 GATTAGATAAATCTCTTTTGGATTTATTAATCAGAAATTTTGCCCTTTTCTTCAAAATAT 367
QY 5463 TACATATGATGATTTATATATAT 5484
DB 368 AATTTAATATATATATATATAT 389

RESULT 15
ABL28217
ID ABL28217 standard; DNA; 4068 BP.
XX
AC ABL28217;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36124.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
XX

CC New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions.
XX
PS Claim 1; SEQ ID NO 36124; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4068 BP; 1067 A; 1021 C; 1110 G; 870 T; 0 U; 0 Other;
Query Match 2.9%; Score 191.8; DB 4; Length 4068;
Best Local Similarity 50.0%; Pred. No. 1.1e-27;
Matches 620; Conservative 0; Mismatches 577; Indels 42; Gaps 4;
QY 3653 CAATATTACCTTACTCAGAGTCTTCTCAGTAGCAGAGATCAGGAGGATTCCTGTTTAT 3712
DB 2817 CACCAGTCCGCGAAATCCCAATTTCTGGGTTCCCGGAGCATGGAGGATTTATCTTTAT 2876
QY 3713 TACACCTACTTTTCAGAACTTGATGATCTCTCATTAACCAAGTAATCCCTTTCTTTGTGG 3772
DB 2877 TCGGCCAAAGTACAGGTGCTTGAGGGAGTATTTGACAGACAAATCCGTAATCTGGTTGG 2936
QY 3773 AATCTTTATCCAGAGCTTCAGATTCCTCGGCAAGGTTTTTCCATAGCGTTTAAATGTT 3832
DB 2937 CGTCTCATCCATAGACACGAGGTACCTTGGGCTAAGGTGCTACCTTTGGGTTGATCCT 2996
QY 3833 GAGATTGGGTGCAGAAATTAAGCATATCTCTCTTAACAGCATCAGAGGCGGAAA 3892
DB 2997 CCGCTTGGGTGCCAGATCGTACTATCCCTGTCCATTGATCTCGGTGCGCGAGGGA 3056
QY 3893 ACCCTTTTGGAGAAATAGGACACACTATTATGAACTTACTTTGTTGACCTTCGAAATTA 3952
DB 3057 GTCCGTGTATGGCGAAATAGTCTAACTATCATTTAACTTTCTGGTGGATTTTCGCAACTA 3116
QY 3953 CCAGTATACCTTGCATTAATATAGATCAACTGTGTGATTCATATGGAATGGGAAAGCTG 4012
DB 3117 TTCGTACTCTCTACCGGCCATCCGAGGATTAATATCCATGGAGGATAGGCGAGCGAC 3176
QY 4013 CATAAATAATACCGGAAAGAGTACAGTGTATGTAATGAAAGTACTAAATTTCTTCCAATGA 4072
DB 3177 GGTGATTAATCCAAAGTACCGCCAGCAGGATGTAATTAAGGCATCAATATGCTAGTGA 3236
QY 4073 GCATGTCTATTAGCAATTTGGAGCAATTTTCAGTACAGAGAGATTCCTCATCTAGTCTGTAT 4132
DB 3237 TCATATACTGGCTTTTGTGCGCAACTTCTCAAAAGTGGCGGATGGTCACTTGGTTTGCAT 3296
QY 4133 ACAGANTGATGGAATTTATGAA-----ACACAGGCCCAACAGTGC 4171
DB 3297 GCAGATATCAACGACGATCGATCCGAGATGTACTCTTCCACGCGGCGGATTAATAT 3356
QY 4172 CACTGGCCCATCTAGAAAGTGACAGGTGCAAGTTTGTGTAATTCATGGAGCTCTAAA 4231
DB 3357 TCAGGTCACCCAGGAAATCACTGGAGCTAGCTTCTTTGTGCTAAATGAAGCCCTCAA 3416
QY 4232 AACATCTTCAGATTTCTTGTAGTCCAGCATAGTTGAAGATGCGCTTAATGTTACAAAT 4291
DB 3417 GAGCAGCAGTGGCTTGTGGGTAAGTCAGCATCTGGAGGATGGCCTAATGGTACAGAT 3476
QY 4292 AACTCCAGAGACCATGAATGGCTTGGGCTAGCTTTACGAGAACAGAAAGACTTTAAAT 4351
DB 3477 TATGCTGCCAAATGGAGGAGTGGCTGAGCGCTTGGCAACCAAAACGACATCGACAT 3536
QY 4352 TACATGTGGGAAAGTTGATG---CAGTAGAAGTCCAGAGAAATACGTTGATATCTCTGGGT 4408
DB 3537 CGTCTGTGGAACCATTTGACGCCACCGACGATCAGAGTGAGATTTGTAAGCATTAAGTGGGT 3596
QY 4409 AGATGCTCAAGAAAGAAACAGAGGATTTACGTTTCAGTGGATGAATATCATTTACA 4468
DB 3597 GGACAAATATCGTGACATCAATGTGGGGGTTAAGTCGCCCAATGATGACGAGCCCATGGA 3656
QY 4469 AGGATTTCCAAAGTGAAAAA---TAAAGTGGAGAGCAGATTTTGAACCCGATGAGAAGAT 4525
DB 3657 TGGCAATTTCAACACATGCGAGTGTGAATGGCCAGCTTTAACTATTCCAAATACCAACTA 3716
QY 4526 TGTAATAATGTACCGAGGTGTTCTACTTTCTTAAAGACCCAGGATTTATCTATTTATCAAC 4585

DB 3717 TGCCATCCGCTGAGTAGATATCTACATTTAATGTCGAGACTGCTATGCAACAATGG 3776
QY 4586 TTCTTTAT-----CAGTTTGCAGAAAGAAATAGCCATGGCTTGTAGTGTGC 4630
DB 3777 TGGAAACTGTGAGGAGATAACACGGATTTCGCGAGCAACTGGCTCGCAGTGCCTCGATGGC 3836
QY 4631 GCTGTGCCCTCACCTGAAAACCTTAAAGAGTAATGGGATGAATAAATTTGGACTCAGAGT 4690
DB 3837 TCTGTGCCACTTCTTCGATCTTTTAGCAGCTGTGGCAATTAATAGCTTTGGCCTAAGAGC 3896
QY 4691 TTCCATTGACACTGATATGTTGATTTTCAGGCGAGATCTGAAGGCCAACTTTCTGSCCTCA 4750
DB 3897 CACGCTCGATCAGGATAACGTGGGCTATGAAGCCGAGCTCGTGGCTCCAAGCTACCGCC 3956
QY 4751 GCATTAATCTAAATGATCTTGTATGCTCTGTATGCTCTGTATGCTGTATGCTGGACCTCCAA 4810
DB 3957 GTGTACATGAACGGGCTGGCAACCACTGTATAGCCACGCTGCATGCGGAATCCTCCGG 4016
QY 4811 CTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTTCAT 4849
DB 4017 CATTCAAGAGCCCATCATTTCTAGAGCTGATCTTTTACAT 4055

Search completed: May 17, 2004, 18:15:14
Job time : 1534.9 secs

This Page Blank (uspto)

REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED REFERENCE	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
REFERENCE AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL REFERENCE AUTHORS	Nature 409, 685-690 (2001) 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002) 6 (bases 1 to 3164) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, O., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, K., Takahashi, F., Takaku-Akahira, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunko Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge. please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. 3164 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM_DB:B130024H06" /db_xref="MGI:2410439" /db_xref="taxon:10090" /clone="B130024H06" /tissue_type="parthenogenote" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="9.5 days embryo" 1. 3164 misc feature /note="hypothetical protein (evidence: ProCrest)"
FEATURES source	1. 3164
ORIGIN	3676 TTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACACCTACTTTTCAGAAA--CT 3733
	Query Match 25.6%; Score 1695; DB 11; Length 3164; Best Local Similarity 78.8%; Pred. No. 0; Mismatches 500; Indels 117; Gaps 16; Matches 2272; Conservative 0
	QY 2663 TAAAGCTCAGGCATTTGAAAGAGTATGATGTCACAACTGGTTCTTAATCTTAACTTCAATCA 2722 Db 379 TATAGCTCAGGCATTTGAAAGAGTATGATGTCAGGTGCTTCTTGTTCTTAACTTCAATCA 438 QY 2723 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2782 Db 439 TTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498 QY 2783 ACCAGCAACTTTTGCACAGTCTCAGCACTTAAACCAACAGGTGTTGAAGGACTATGTTCCAA 2842 Db 499 ACCTACCACTTTTGCACAGTCTCAGCACTTAAACCAACAGGTGTTGAAGGACTATGTTCCAA 558 QY 2843 AGAACAGAGAGAGTATGTTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2902 Db 559 AGAACAGAGAGAGTATGTTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 618 QY 2903 ACAAATTTATCATCTCGGAAGTAAAGATGTTCTGATGACCTTTAGTTCCTGCTTCTTACC 2959 Db 619 TAGCAATTTATCATCTCGGAAGTAAAGATGTTCTGATGACCTTTAGTTCCTGCTTCTTACC 678 QY 2960 TGATGTGCTTATGACAGTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAAAGCC 3019 Db 679 TGATGTGCTTATGACAGTAAACACAGTGGATCATTTCCCATTTCTACTACAGTGGAAAAAGCC 738 QY 3020 AAACATGACAGAGAGTATTTACAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGAT 3079 Db 739 AAACATGATTTAGGAGACATTTATAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 798 QY 3080 TCCATCAGTGGAAAAATTTCTATGAACACAGCAAGTAAAGGAGTAAAGGAGTAAAGGAGTAAAGGAG 3135 Db 799 TGCACCTGTGGAAGATGCGCGGAGACACAGCGACGAGAGGCTTTGGCCCATGSCCTGG 858 QY 3136 GTTCATTACACTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3195 Db 859 CCCCCTTTCACACTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918 QY 3196 GTGCTCTAGTTAAACAGCAATTTACCTATTGCTAGTATTTTTCAGAACTGAAGACATCTAGTCTCTCTG 3255 Db 919 TTGTCCTCTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 978 QY 3256 ATATTAAACAGTATGCTGCAATAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3315 Db 979 GTGTGCAAAATGCTGTGCAACAACTAGTCTTCTACAGACGATGACATGATGATGATGATGATGATGAT 1038 QY 3316 CCCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3375 Db 1039 CTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1098 QY 3376 CTCTAGCAGATCATTTTGTCTTCTTGAAGGTGAAGGCTTTTCACTCTGTTTCACTTTGTGCC 3435 Db 1099 CTCTAGCAGATCATTTTGTCTTCTTGAAGGTGAAGGCTTTTCACTCTGTTTCACTTTGTGCC 1158 QY 3436 TAAATGCTTAATCTCTGTAATGTCATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 3495 Db 1159 TAAATGCAAACTCTCTGTAATGTCATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1218 QY 3496 ACTTTTCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3555 Db 1219 ACTTCTCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1278 QY 3556 GTTTCGCAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3615 Db 1279 GTTTCGCAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338 QY 3616 AGGATGCTCTTAAAGGAAAAATACATAGAAAACTTGGACATATTACCTTTTACTGAGAGTT 3675 Db 1339 AGGATGCTCTTAAAGGAAAAATACATAGAAAACTTGGACATATTACCTTTTACTGAGAGTT 3733

```
Db 1367 -----GGGGTTCCTTGGTTTATATACACTACCTTCCAGAAACCTT 1406
Qy 3734 TGTATCTCTCATTAACCAAGTAATCCTTTTCTTTGTGGAAATCTTATCCAGAAGCTTGA 3793
Db 1407 TGTATGCTCTCCCGTGCACAGGAGTCTTTTCTCTCGCGCATCTTATCCAGAAGCTGA 1466
Qy 3794 GATTCCTCTGGG-CAAAAGGTTTTCTTATGCTTAAATGTTAGATTTGGGTGCAAGATATA 3852
Db 1467 GATTCCTCTGGGCAAAAGGTTTTCTTATGCGGTAAATGTTGCGATTTGGGAGCAAGATATA 1526
Qy 3853 AAGCATATCTCTGCTCTCTAA-CAAGCATCAGAGGC-CGAAACCTCTTTTGGAGAAAT 3910
Db 1527 AAGCATACCTCTCTCTAAACCAAGTGTGAGAGGAGGAAACCTCTTTTGGAGAAAT 1586
Qy 3911 AGGACACACTATTATGAACTTACTTGTGACCTTGGAAATACAGATACCTTGTGATTA 3970
Db 1587 AGGACATACACTATTATGAACTTACTTGTGATCTTGGAAATACAGATACCTTGTGATTA 1646
Qy 3971 TATAGATCAACTGTTGATTCATATGGAATGGGAAAGCTGCATATAAATACACGGAA 4030
Db 1647 CATCGATCAGCTATTGATTCATATGGAATGGGAAAGCTGTATAAATACCTCGAA 1706
Qy 4031 AAGTACAGTGTATGTAATGAAAGTACTTAAATCTTCCAATGAGCATGTCAATGACATGG 4090
Db 1707 AAGTACAGTGTATGTAATGAAAGTACTTAAATCTTCCAATGAGCATGTCAATGACATGG 1766
Qy 4091 AGCAAGTTTCAGTACAGAGAGATCTCATCTAGTCTGTATACAGATGATGAAATTA 4150
Db 1767 AGCTAGTTTTAGTACAGAGAGAGATCTCATCTGTGCTGTGTACAGAGTATGGAGTTTA 1826
Qy 4151 TGAACACAGGCCAACAGTGCCTGCGCCATCTTAGAAGAAAGTGACAGGTGCAAGTTTGT 4210
Db 1827 TCAGACTCAGGCCAACAGTGCCTGCTGACCAAGAAAGTGACAGGTGCAAGTTTGT 1886
Qy 4211 GGTATTCAGTGGAGCTCTAAACATCTTCAGATTTCTTGTCTTAAAGTCCAGATAGTTGA 4270
Db 1887 GGTGTTTAAATGGTCTCTGAAACATCTTCAGGATTTCTTGCCAAATCCAGCATAGTGA 1946
Qy 4271 AGATGGCTTAAATGTCACAAATACTCCAGAGACCATGAATGGCTTGGCGCTAGCTTTACG 4330
Db 1947 AGATGGCTTAAATGTCACAAATACTCCAGAGACCATGGAGGCTTGGCGCTAGCTTTACG 2006
Qy 4331 AGAACAGAAAGACTTTAAGATCCAGTGTGGAAGAAAGTTGATCGGTAGACCTGAGAGAT 2066
Qy 4391 CGTGGATCTGCTGGCTAGATGCTGAGAAAGAAAGAAACAAAGAGTTATCAGTTCCAGT 4450
Db 2067 TGTGGACATCTGTTGGGTGGATTGAGAGAGAGAGAAACAAAGGTGTTATCAGTTCCAGT 2126
Qy 4451 GGATGGAAATCATTAACAAGGATTTCCAAGTGAAGAAATATAAATGGAAGCAGATTTGA 4510
Db 2127 AGATGGAAATGTCAGTGGAGGATTTCCAAGTGAAGAAATATAAATGGAAGCAGATTTGA 2186
Qy 4511 AACCGATGAGAAAGTTGTAATGATACCGAGGTTTCTACTTCTTAAAGCAGGATTTT 4570
Db 2187 AACTGAGAGAAAGTGTGAAATGCTAGAGTGTCTTACTTCTTAAAGCAGGATTTT 2246
Qy 4571 ATCTATTATCAACTTCTTATCAGTTTGAAGAAATAGCCATGGCTGTAGTGTCTGC 4630
Db 2247 ATCTATTATCAAGTAGTACAGTTTGAAGAGAGATGCTGTGCTGTAGTGTCTGC 2306
Qy 4631 GCTGTGCCCTCACCTGAAACCTTAAAGTAATGGAATGGAATGGAATGGAATGGAATG 4690
Db 2307 CCTGTGCCCTCACCTGAAACCTTAAAGTAATGGAATGGAATGGAATGGAATGGAATG 2366
Qy 4691 TTCCATTGACACTGATGTTGATTTTCCAGCAGGATCTGAAGCCAACTTCTGCTCA 4750
Db 2367 GTCCATTGACACTGATGTTGATTTTCCAGCAGGCTGTGAAGCCAACTTCTGCTCA 2426
Qy 4751 GCATTATCTAAATGATCTTGTATGCTCTGATACCTGTGATCCATGCTGGGA-CTTCCA 4809
```

```
Db 2427 GCATTACCTGAATGATCTGGATAGTGCCCTGATCCCTGTGATCCACGGCGGACCCCTCA 2486
Qy 4810 ACTCTAGTTTACCATTAGAAATAGAAATAGTGTGTTTTCATTATAGAAATCTTTTGTAGT 4869
Db 2487 ACTCTAGTCTGCCACTAGAAATAAGATTTTTCATTCTAGAAAACCTTTCTGAGT 2546
Qy 4870 GAAAGAAATGTCATATACATATTTGCAACCTAAATTTGTTAAACCTAACTCCAGCACTAA 4929
Db 2547 GAAACCGTGTACTGTTGGTGTATTTTGCAAACTGATGGGTAAACGAAATCCAGCACTAA 2606
Qy 4930 AGCTGAAATGTCACAAACACATTAAGATTAATATGCTGATTTTGAACACATAGCT 4989
Db 2607 AACGGAATGTCACAAACACATAA-----ATGTTTGATAATTTGAATATATAATTT 2656
Qy 4990 TTGCTCTTTAGCAGGAATGATCTTTTCAAACTATTAGCAACAATTTTAAATATCTAAAA 5049
Db 2657 TCTTTTATAGTAGAGAAATGATTTTAAATATCATTAGCAGATTTTAAAGTCTCTAAAA 2716
Qy 5050 ATTTAAGATGATCCTATCTTCTGATGCTTCAAAATTAATTAAGTACTAAAAAGCAAG 5109
Db 2717 ATTTAAG-GAATCATCAATTTTATAGCTTATGCTTATTAAGAACTAAAGAGG--AAAT 2773
Qy 5110 ATTTCTTTTAAAGAAATTTATGATGCTTCTGATGCTTCAAAATTAATTAAGTACTAAAAAGCAAG 5169
Db 2774 ATTTCTTAACTTAAATTTTACAGTATTTTACTGTGTACCTTAACTGTTAAAGCCAAAGCACT 2833
Qy 5170 GCACCTTAGTATACCTCTTTTATGCCAATAATGATTTTAAATGAAGCTCTTTTTCAGATGTA 5229
Db 2834 GCGCTAGCCCTACTCTCTCATAGATGACGA----- 2863
Qy 5230 ACCTTATGAGAAATATCTGCTTTGTGTATATGCGCAGTATAGAACTAGTGGTTCTAAAGT 5289
Db 2864 ---GAATGAAGAGAGAGACCTGCGTTGTGTGGATGTAGTGAGAACACTGGTTT-TGAAGT 2919
Qy 5290 CTGTCAAAATGTATTTTCACTGGCACAAAACCCAGTTTGTAGGTCTTAGACTTATAATCT 5349
Db 2920 CTGTTTAAATGTATCTCAGTAAAGAGACAGTTTCAAGATCTTAGACTTATAATCT 2979
Qy 5350 TTGAATAAACTGATAACTTATTGTATTAATTTGGAGTGGAGACCTACCTCCATATATGA 5409
Db 2980 TTGAATAAACTGATAACTTATTGTATAAATGGAATGAAGACCTACCTCCATGATTACA 3039
Qy 5410 TAACTCTTTTGGATTATAATCAGAAATTTTCCCTTTTCTTCTCAAAATTTACATAT 5469
Db 3040 TACAACTGTTT-GATTAAATTTAGAAATTTTCCCTTTTATTTCTTCAAAATTTCAATAT 3098
Qy 5470 GTATGTTATATATATCCACATATATAGTTTCCCTGATTAATGGAATATAAAATAATG 5529
Db 3099 CTATATA-TATATATACACATATAAATATATATCTTTTCTG---ATTAAAAATAATG 3154
Qy 5530 CCGGTGCTT 5538
Db 3155 CCGGTGCTT 3163
```

RESULT 2
BM925782
LOCUS
DEFINITION
AGENCY 6710339 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5764025
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM925782 1035 bp mRNA linear EST 12-VAR-2002
AGENCY 6710339 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5764025
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM925782.1 GI:19376161
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://img.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM12816 row: m column: 18
 High quality sequence stop: 794.
 Location/Qualifiers
 1..1035
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5764025"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_114"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

FEATURES

source

ORIGIN

Query Match 14.4%; Score 951.8; DB 12; Length 1035;
 Best Local Similarity 96.8%; Pred No. 7.8e-179;
 Matches 1003; Conservative 0; Mismatches 28; Indels 5; Gaps 3;

QY 5277 TGGTTCTTAAGCTCTGCAAAATGTTATTCAGTGGCAAAAACACAGTTTGGAGTCTTA 5336
 Db 1 TGGTTCTTAAGCTCTGCAAAATGTTATTCAGTGGCAAAAACACAGTTTGGAGTCTTA 60

QY 5337 GACTTATATCTTTTGAATAAACTGATACTTATTTGTATAATTTGAGTGGAGACCTAC 5396
 Db 61 GACTTATATCTTTTGAATAAACTGATACTTATTTGTATAATTTGAGTGGAGACCTAC 120

QY 5397 CTCATAATTAGATAAACTCTTTTGGATTATATCAAGATTTTGCCTTTTCTTCTCA 5456
 Db 121 CTCATAATTAGATAAACTCTTTTGGATTATATCAAGATTTTGCCTTTTCTTCTCA 180

QY 5457 AATTATTACATATGATGATATATATATCCACATATAGTTTCCCTGATTAAATCGAT 5516
 Db 181 AATTATTACATATGATGATATATATATCCACATATAGTTTCCCTGATTAAATCGAT 240

QY 5517 ATTAAATAATTGGGGTGGCTTCAGGACTTTTTCCTCTATATTTAAGTATATGTTTT 5576
 Db 241 ATTAAATAATTGGGGTGGCTTCAGGACTTTTTCCTCTATATTTAAGTATATGTTTT 300

QY 5577 ATAGCAGAACATATTCGAATGTTTATTAATCTTTTAATTAATTTATGATAGTATAT 5636
 Db 301 ATAGCAGAACATATTCGAATGTTTATTAATCTTTTAATTAATTTATGATAGTATAT 360

QY 5637 TTTTGTATCAAGTGCATATTTTTCCTTCCTTCCTCCAACTATACCACCTGATTT 5696
 Db 361 TTTTGTATCAAGTGCATATTTTTCCTTCCTTCCTCCAACTATACCACCTGATTT 419

QY 5697 TACCATTCTTAAGTAGTACTGACGACGGGCCAGATGACCTTGAAGTAGTCATATGATAG 5756
 Db 420 TACCATTCTTAAGTAGTACTGACGACGGGCCAGATGACCTTGAAGTAGTCATATGATAG 479

QY 5757 CAATAAATGAGCTGGAACAGGTTTTTTTACTTCCACTTTAATCCCTTAGAATTTCTTG 5816
 Db 480 CAATAAATGAGCTGGAACAGGTTTTTTTACTTCCACTTTAATCCCTTAGAATTTCTTG 539

QY 5817 GCAACTTCGCATATTTTCAATTCACACGGTGTATAGTATAAAATTTAAATGAACATTA 5876
 Db 540 GCAACTTCGCATATTTTCAATTCACACGGTGTATAGTATAAAATTTAAATGAACATTA 599

QY 5877 CTTTGGCATATTTTAAATCTTATATGTTAGTATTTTTTTTATAACAGGATATTACATA 5936

Db 600 CTTTTCATATTTTAAATCTTTTATATAGTACTTATTTTATAACAGGATATTACATA 659

QY 5937 AGTTAATCCTATGTTTGAATTTGTAATGTTTACAGAGCTTTCTCTTTTACTTCAACAGCAA 5996

Db 660 AGTTAATCCTATGTTTGAATTTGTAATGTTTACAGAGCTTTCTCTTTTACTTCAACAGCAA 719

QY 5997 AAAGTGGGGGCATATTTGATGCTCTGTCATTTAAGTTATGTAAAAATTTAATCATTT 6056

Db 720 AAAGTGGGGGCATATTTGATGCTCTGTCATTTAAGTTATGTAAAAATTTAATCATTT 779

QY 6057 TTGATGCTTTAAACATTTCTCATGTTGTAATATATGTTTTTGTATCAAAACACTCATAT 6116

Db 780 TTGATGCTTTAAACATTTCTCATGTTGTAATATATGTTTTTGTATCAAAACACTCATAT 839

QY 6117 TTCAAGAAAAAGAAATTTATGTTAAATAGCCCTGTTTAAAGAAAAATTTTATGAAGCATC 6176

Db 840 TTTCAGAAAAAGAAATTTATGTTAAATAGCCCTGTTTAAAGAAAAATTTTATGAAGCATC 899

QY 6177 TCAACTTGAAGTCAAGTCAAAAGTTATTAACCTCAGGATCTGAGGTTCTCAAGCTAGGAGAGA 6236

Db 900 TCACCTTGAAGTCAAGTCAAAAGTTATTAACCTCAGGATCTGAGGTTCTCAAGCTAGGAGAGA 959

QY 6237 CTGAGAAATTTAATCAGTTTGGGCATA---TAGTTTGGACTGAATCAGAT-CTGTAGTAC 6292

Db 960 CTGAGAAATTTTAAATCAGTTTGGGCATATTTAGTTTGGACTGAATCAGATCTCTGGAGAAC 1019

QY 6293 TTAGCCAAAGACAATT 6308

Db 1020 TTAGCCAAAGACAATT 1035

RESULT 3
 AK045090
 LOCUS
 DEFINITION
 AK045090 1973 bp mRNA linear HTC 20-SEP-2003
 Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone:B13031L15 product:weakly similar to ENDOFIN [Homo sapiens], full insert sequence.

ACCESSION
 AK045090
 VERSION
 AK045090.1 GI:26337062
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2
 10349636

AUTHORS
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
 MEDLINE
 PUBMED
 20499374
 11042159

REFERENCE
 3
 11042159

AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishire, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
 MEDLINE
 PUBMED
 20530913
 11076861

REFERENCE
 4

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M13211 row: i column: 07
High quality sequence stop: 656.
Location/Qualifiers
1. .1120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6015438"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

FEATURES
source

ORIGIN

Query Match 12.1%; Score 801.8; DB 13; Length 1120;
Best Local Similarity 91.0%; Pred. No. 5.4e-149;
Matches 954; Conservative 0; Mismatches 78; Indels 16; Gaps 9;

QY 1219 TCAATATTACCTGACCTTTCCCTTCAGGAGATAAGACT-----GTTATTAACAACATCTG 1272
DB 1171 TCAAGTTACCTGACCTTTCCCTTCAGGAGATAAGACTTTGTGTTTCATGAACAGACTG 1230
QY 1273 CACAGAAGAGCTCAAAAAGTTTACAGCTTAAGGATATGATGTAATCAAGATTCCTCTT 1332
DB 1231 TAAAGAAGAGCTCAAGAATTTAGACCTTAAGATATTAATGATAT----- 1276
QY 1333 CAGCTTTACATGTTTCCAGTAAAGATGCGCGTCTCATTTCTCTGCTCTCTCGCTG 1392
DB 1277 ---AGTCCATGTTTTCAGGTGATGATGTCGCCCTTCACTGCTCTGCTTTTCATGTCG 1332
QY 1393 GGTCTATGCTGGATCATTAATTAAGAGTAAAGACAGCGGGTGATTTTACCTCAGCATG 1452
DB 1333 GGTCTTTGTGATGCTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1392
QY 1453 AACATAAAGATTAATATAAAGATGAGTACTATACATGAAGAATAACAGAACAGTGTG 1512
DB 1393 AGTCTGAAGGCGAGAACATGATGATGATGATGATGATGATGATGATGATGATGATG 1452
QY 1513 TTCTAGTGGGGAACCATCAAGAGATGATCTTTTGAACAGGAAATGTAAGACA 1572
DB 1453 TTTTAGATGGG-----GAGACAGATCTTTCAAGAGAGAACGTTAGGAGCA 1500
QY 1573 TACTCTTCAGTCATTAATTAAGAGGATGGAAGACAGAAAGATAGATCTTGACACAGAC 1632
DB 1501 TATTTCTTCAGCCAGTAATGAAGAAGAGGAGGAGGAAAGGTTGAGGTTGAGGATGG 1560
QY 1633 TAATCAGAGCTGAGTCTTTTGA-----TGTTGGTGACACAGATTCACAGTTGAGAA 1686
DB 1561 TAATCAGTGGTGAGTCTTTGAGATGCTTCCAGAGGCTGCTGCTGACAGCTGGT 1620
QY 1687 CTCAGAGGGGCTTCTGCACTATGTCAGAGTCTTCTGATGTTGTTGAGGTTTGA 1746
DB 1621 CTCGGTGGCATTTCTGCTCAAGTGTCCAGAGGCTCTGCTGCTGCTGCTGCTGCTGCT 1678
QY 1747 TTAATACCTTTTCAAGCAATGATATGATGAGGCAAGCTTAGATTAATTAATGATG 1806
DB 1679 ---CACCTTTCCAGCAGTATGATGATGATGATGATGATGATGATGATGATGATG 1734
QY 1807 AAGGCGAAAGTGGCCCACTAATAGTATGATGATGATGATGATGATGATGATGATGATG 1866
DB 1735 AAGCATGAGAGTGGCATACTAATAGCATGCTGAACTGATGCTGATGCTGATGCTGAT 1794
QY 1867 AGTATCTTCAGACCACTAACATAGTCTTTTGAAGAAATGTAATGACACTTAAATGCG 1926
DB 1795 AGTGCTTTTCCAGCTTAACACATGCTGCTGGAGAAATGTAATGATGATGATGATGAT 1854
QY 1927 AATGAATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1986
DB 1855 AGATGAATCAGATCATACTATGAAGGACTACATGATGAATAATGCTGGGATATATCT 1914
QY 1987 ATGCAAGAGCAGGCTATTGGGAAAGTCAATGATGATGATGATGATGATGATGATGAT 2024
DB 1915 ATGCTGAAGCAGGAGCTGCTGGAGAAATGAGAGTGT 1952

RESULT 4

BQ430548
LOCUS BQ430548 1120 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT 7766657 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6015438
5', mRNA sequence.
ACCESSION BQ430548
VERSION BQ430548.1 GI:21169624
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1120)
AUTHORS NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Db 657 CCATGATAGTACTGTACAGAACACAGATGATACAGTTCTGAATTACAAAATAG 716
 QY 845 AGAATCGGAGGAATCAAGAAATGGGTATATAAAGTAGATACACAC-TTTCAGATTCT 903
 Db 717 AGAATCGGAGGAATCAAGAAATGGGTATATAAAGTAGATACACACTTTTCAGATTCT 776
 QY 904 ATAAATACAGTGAACAG-AAAATTTAAAGATAAAAAGATCTTTTAATCAGTTAGATCA 962
 Db 777 ATAAATACAGTGAACAGAAAATTTAAAGATAAAAAGATCTTTTAATCAGTTAGATCA 836
 QY 963 ATTGTTGATTTAAACATGTCATCTGTTTACCTCGACCAAGTTCCAAAATGTTTCATG-C 1021
 Db 837 TTTGTTGATTTTACATGTCATCTGTTTGAATCCAAAAGTTCCAAAATGTTTCATGCC 896
 QY 1022 CAAAGACAAGCTACACACAG-AGCCAGCCATGT-GGATTAATAAAGATGTTG- --- 1075
 Db 897 CAAAGACCAGCTACACCCAGAACCCAGCCAGCATGTGGGATTAATAAAGATGTTGGCTT 956
 QY 1076 -CTTAGTAAAGAGGAGTAGATGTGGCAGTCATACTGCGG-CAGAAATGTTTAAAGAA 1133
 Db 957 AATTAAAGGAGGAGTATTATGTGGCCATTCCTTACCTGCGCCAAAATGTTTAAAGAAA 1016
 QY 1134 GAGGG- ---CAAGACAAGTCTTTGACCTGCAGCCCTTCCGAAAAATCAAG-ATTATGCT 1188
 Db 1017 GAAGGCAAGACAGGTGCTTTTAACCTGCGACCTTCCGAAAAGGAGAAATATGCG 1076
 QY 1189 TAAATGATTCAAATTCAGAGATGAATA 1216
 Db 1077 TTAAGGGTCCAAATCCAGAAAAGAAA 1104

RESULT 5

BI767442
 LOCUS 892 bp mRNA linear EST 25-SEP-2001
 DEFINITION 603057595F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207230 5',
 mRNA sequence.
 ACCESSION BI767442
 VERSION BI767442.1 GI:15759020
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

1 (bases 1 to 892)
 NIH-MGC <http://mgs.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM1519 row: m column: 23
 High quality sequence stop: 788.
 Location/Qualifiers

FEATURES
source

1..892
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5207230"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_122"
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size

range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 11.5%; Score 759.8; DB 12; Length 882;
 Best Local Similarity 95.6%; Pred. No. 1.2e-140;
 Matches 825; Conservative 0; Mismatches 32; Indels 6; Gaps 4;
 QY 471.6 TTTCAGGAGGATCTGAAGGCCAACTCTGCTCAGCAATATCTAAATGATCTTGATAGT 4775
 Db 1 TTTCAGGAGGATCTGAAGGCCAACTCTGCTCAGCAATATCTAAATGATCTTGATAGT 60
 QY 4776 GCTCTGATACCTGTGATCCATGCTGGGACCTCCAACTCTAGTTTACCAATGAAATAGAA 4835
 Db 61 GCTCTGATACCTGTGATCCATGCTGGGACCTCCAACTCTAGTTTACCAATGAAATAGAA 120
 QY 4836 TTAGTGTTTTTCAATATAGAACATCTTTTGTGTAAGAGATGTCATATACATATG 4895
 Db 121 TTAGTGTTTTTCAATATAGAACATCTTTTGTGTAAGAGATGTCATATACATATG 180
 QY 4896 CAACCTAAATTTTAAAACTAACTCCAGCACCTAAAGCTGAAATGCCACAAACACTAAAG 4955
 Db 181 CAACCTAAATTTTAAAACTAACTCCAGCACCTAAAGCTGAAATGCCACAAACACTAAAG 240
 QY 4956 TATAAATATGCTGATTTTTTGAACACATAGCTTGTCTTTTAGGAGGAGATGATCTTT 5015
 Db 241 TATAAATATGCTGATTTTTTGAACACATAGC-TTGCTCTTTTAGGAGGAGATGATCTTT 299
 QY 5016 TCAATCATTAGACCAATATTTAAATATCTAAATTTAAAGAGATCCATCTCTCTGTAG 5075
 Db 300 TCAATCATTAGACCAATATTTAAATATCTAAATTTAAAGAGATCCATCTCTCTGTAG 359
 QY 5076 CTTTCAATATTAATTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTTATAGCAAT 5135
 Db 360 CTTTCAATATTAATTAAGTACTAAAGACAGAGATTTCTTTTAAAGAAATTTATAGCAAT 419
 QY 5136 TACTGTGTTATTTAAATGCTTAAGCCAAAGTATCTGCACCTTAGGTATACCTCTTTATGCCA 5195
 Db 420 TACTGTGTTATTTAAATGCTTAAGCCAAAGTATCTGCACCTTAGGTATACCTCTTTATGCCA 479
 QY 5196 ATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAAATATCTGCTTTG 5255
 Db 480 ATAATGATTTTAATGAAGGCTCTTTTCAGATGTAACTTATGAAGGAAATATCTGCTTTG 539
 QY 5256 TGTATATGCCAGTTAGATAGTCTGTTTCTAAAGCTGTCAAAATGTTATTCAGTGGCACA 5315
 Db 540 TGTATATGCCAGTTAGATAGTCTGTTTCTAAAGCTGTCAAAATGTTATTCAGTGGCACA 599
 QY 5316 AAAACCACTTTTGAGGCTCTTAGACCTTATAATTTCTTGAATAAACTGATAACTTATTTGT 5375
 Db 600 AAAACCACTTTTGAGGCTCTTAGACCTTATAATTTCTTGAATAAACTGATAACTTATTTGT 659
 QY 5376 ATAATGAGTGGAGACCTACCTCCATAA-TTAGATAAATCTTTTGGATATTAATCAG 5434
 Db 660 ATAATGAGTGGAGACCTACCTCCATAAATTTAGATTAAACTCTTTTGGATATTAATCAG 719
 QY 5435 AATTTTGCTCTTTTCTTCTCAAAATTTTACATATGATGATATATATATCCACATATAT 5494
 Db 720 AATTTTGCTCTTTTCTTCTCAAAATTTTACATATGATGATGATGATGATGATGATGAT 778
 QY 5495 AGTTTTCCTGATTAATAGGATATTAATAATTAATTCGGGGTCTTCAGACATTTTGTCTTC 5554
 Db 779 AGTTTTCCTGATTAATTAATTCGGGGTCTTCACATAAATTCGGGGTCTTCGGGACCTTTTGCCTCA 838
 QY 5555 TATATTTAAGTATATGTTTATA 5577
 Db 839 T---TTTAAAGTATATTTGTATATA 858

RESULT 6
 BQ431811

```

LOCUS      BQ431811      836 bp      mRNA      linear      EST 24-MAY-2002
DEFINITION AGENCOURT_7907398 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6154816
5', mRNA sequence.
ACCESSION  BQ431811
VERSION    BQ431811.1 GI:21170887
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1. (bases 1 to 836)
            NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13496 row: h column: 17
            High quality sequence stop: 508.
            Location/Qualifiers
                1..836
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6154816"
                /tissue_type="retinoblastoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH-MGC_67"
                /note="Organ: eye; Vector: pCMV-SPORT6; Site1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 1.75 kb. Library constructed by Life
                Technologies."

ORIGIN
Query Match      11.3%; Score 747.6; DB 13; Length 836;
Best Local Similarity 96.7%; Pred. No. 3.3e-138;
Matches 784; Conservative 0; Mismatches 25; Indels 2; Gaps 2;

QY 3127 CTACTTCTGGTTCATTACACTAGATGATGATGTTTTCAGAACTGAAGAACCATCTA 3186
DB 1 CTACTTCTGGTTCATTACACTAGATGATGATGTTTTCAGAACTGAAGAACCATCTA 60

QY 3187 GTCCTACTGGTGTCTTACCTAAGCAATTTACCTATTGCTAGTATTTCAGATTATAGGT 3246
DB 61 GTCCTACTGGTGTCTTACCTAAGCAATTTACCTATTGCTAGTATTTCAGATTATAGGT 120

QY 3247 TACTGTGTGATATTAAACCAAGTATGTCGCAATAGATTAGTCTTCTACCTAATGATCAGG 3306
DB 121 TACTGTGTGATATTAAACCAAGTATGTCGCAATAGATTAGTCTTCTACCTAATGATCAGG 180

QY 3307 ACAGTTTCCGCCCACTTCTGGTTCATCTGGAGAAAGGGATCAGTCCCTGTAGTAGAAG 3366
DB 181 ACAGTTTCCGCCCACTTCTGGTTCATCTGGAGAAAGGGATCAGTCCCTGTAGTAGAAG 240

QY 3367 AACATCCATCTCATGAGCAGATCATTTTCTTCTTGGAGTGAAGGCTTTCATCCTCTGTA 3426
DB 241 AACATCCATCTCATGAGCAGATCATTTTCTTCTTGGAGTGAAGGCTTTCATCCTCTGTA 300

QY 3427 CATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACA 3486
DB 301 CATTTGTCCTAAATGCTAATCTACTCGTGAATGTCAAATTCATATTTTATTCCTCAGACA 360

QY 3487 AATATTGCTATTTCACCAATGATGATGCTGGGACGCGAGCAATATTATTC 3546
DB 361 AATATTGCTATTTCACCAATGATGATGCTGGGACGCGAGCAATATTATTC 420

QY 3547 TATTGTTATGTTTGCCAAATGAAGACTACTATTTCCTAGGACATCTTCAGACTATTTATCA 3606

```

```

DB 421 TATTGTTAATGTTGCAAAATGAAGACTACTATTCTTAGGACATCTTCAGACTATTTATCA 480
QY 3607 CCATATATAAGGATGCTCTAAAGAAATAACATAGAAAACCTTGGACAAATATTACCTTTA 3666
DB 481 CCATATATAAGGATGCTCTANAGAAATAACATAGAAAACCTTGGACAAATATTACCTTTA 540
QY 3667 CTGACAGATTTTCTCAGTAGCAGGATCAGGAGGATCCCTGTTTATTACACTACTTTTC 3726
DB 541 CTGACAGATTTTCTCAGTAGCAGGATCAGGAGGATCCCTGTTTATTACACTACTTTTC 600
QY 3727 AGAAAATTCATGATCTCTCATTTACCAAGTAATCTCTTTTCTTGGAATTTTATCCAGA 3786
DB 601 AGAAAATTCATGATCTCTCATTTACCAAGTAATCTCTTTTCTTGGAATTTTATCCAGA 660
QY 3787 AGCTTGAGATTCCTGGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATGGGTGCGAG 3846
DB 661 AGCTTGAGATTCCTGGGCAAGGTTTCTCTATGCGTTTAAATGTTGAGATGGGTGCGAA 720
QY 3847 AATATAAGGATPAT-CCTGCTCTCTTAAAGCATCAGAGG-CGAAAACCTCTTTTGTG 3904
DB 721 AATATAAGGATPAT-CCTGCTCTCTTAAAGCATCAGAGG-CGAAAACCTCTTTTGTG 780
QY 3905 AGAAAATAGGACACACTATTATGAACCTTACTT 3935
DB 781 GAGAAAATAGGACCCACTATTATGAACCTT 811

RESULT 7
BQ884268      897 bp      mRNA      linear      EST 16-AUG-2002
LOCUS        AGENCOURT_8682001 Lupski_sciatic_nerve Homo sapiens cDNA clone
DEFINITION  IMAGE:6197558 5', mRNA sequence.
ACCESSION  BQ884268
VERSION    BQ884268.1 GI:22276276
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1. (bases 1 to 897)
            NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13607 row: m column: 15
            High quality sequence stop: 607.
            Location/Qualifiers
                1..897
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6197558"
                /sex="male"
                /tissue_type="sciatic nerve"
                /dev_stage="adult, 70 yr"
                /lab_host="DH10B"
                /clone_lib="Lupski_sciatic_nerve"
                /note="Vector: pCMV-SPORT6 (Life Technologies); Site1:
                NotI; Site 2: SalI; cDNA made by oligo-dT priming.
                Directionally cloned using the following adaptors:
                5'-TCGACCCAGCGCTCCG-3' and
                5'-GACTAGTCTAGATCGGAGCGGCCCT(15)-3'. Size selected >
                1 kb for average insert length 1.87 kb. This is a primary
                library, non-amplified. Library constructed by Life

```

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Query Match	11.2%;	Score 741.8;	DB 13;	Length 897;
Best Local Similarity	94.2%;	Prod. No. 4.7e-137;		
Matches 847;	Conservative	0;	Mismatches 42;	Indels 10; Gaps 7;
QY	4489	TAAAACTGGAAGCAGATTTTCAAAACCGATGAGAAGATTGTAATAATGTACCGAGGTGTTCT	4548	
DB	1	TAAAACTGGAAGCAGATTTTCAAAACCGATGAGAAGATTGTAATAATGTACCGAGGTGTTCT	60	
QY	4549	ACTTTCTAAAGGACAGAGTTATCTATTTTATCAACTTCCTATATCGATTGTGCAAAAGAAA	4608	
DB	61	ACTTTCTAAAGGACAGAGTTATCTATTTTATCAACTTCCTATATCGATTGTGCAAAAGAAA	120	
QY	4609	TAGCCATGGCTTTGATGTCGCGGTGCGCCTCACCTGAAACCTCTAAAAGTAATGGGA	4668	
DB	121	TAGCCATGGCTTTGATGTCGCGGTGCGCCTCACCTGAAACCTCTAAAAGTAATGGGA	180	
QY	4669	TGAATAAAATTGACTCAGAGTTTCCATATGACATATGATATGTTTGAATTTTCAGCAGGAT	4728	
DB	181	TGAATAAAATTGACTCAGAGTTTCCATATGACATATGATATGTTTGAATTTTCAGCAGGAT	240	
QY	4729	CTGAAGGCCAACTTCGCTCAGCATTTATCTTAATGATCTTGATAGTGCCTCGATACCTG	4788	
DB	241	CTGAAGGCCAACTTCGCTCAGCATTTATCTTAATGATCTTGATAGTGCCTCGATACCTG	300	
QY	4789	TGATCCATGTTGGGACCTCCAACTCTAGTTTACCATTTAGAAATAGAAATTAGTGTTTTTCA	4848	
DB	301	TGATCCATGTTGGGACCTCCAACTCTAGTTTACCATTTAGAAATAGAAATTAGTGTTTTTCA	360	
QY	4849	TTATAGAACAATCTTTTTTAGTGAAGAAATGTGCCATTTACATATTTCACACCTAATTTGT	4908	
DB	361	TTATAGAACAATCTTTTTTAGTGAAGAAATGTGCCATTTACATATTTCACACCTAATTTGT	420	
QY	4909	TAAAACTAACTCCAGCACTAAAGCTGAAATGCCACAAACACTAAAAGTATAAATATATGTCT	4968	
DB	421	TAAAACTAACTCCAGCACTAAAGCTGAAATGCCACAAACACTAAAAGTATAAATATATGTCT	480	
QY	4969	GATTTTGAACACATAAGCTTTTGCTCTTTAGCGCAGGAATGATCTTTTCAAAATCATTAGC	5028	
DB	481	GATTTTGAACACATAAGCTTTTGCTCTTTAGCGCAGGAATGATCTTTTCAAAATCATTAGC	540	
QY	5029	ACAATATTTAAATATCTAAAAATTTAAGAGATCCATATCTTCTGTAGCTTTACAAATTAAT	5088	
DB	541	ACAATATTTAAATATCTAAAAATTTAAGAGATCCATATCTTCTGTAGCTTTACAAATTAAT	600	
QY	5089	TTAAGTACTAAAAGACAGAGTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTATTT	5148	
DB	601	TTAAGTACTAAAAGACAGAGTTCTTTTAAAGAAATTTATAGCATTTACTGTGTTATTT	660	
QY	5149	AAATGCTTAAGCCAAAGTATCTGCACCTTAGGTATACCTCTTTTATGCCAATAATGATTTTAA	5208	
DB	661	AAATGCTTAAGCCAAAG-ATCTGCCCTTAGGTATACCTCTTTTATGCCAATAATGATTTTAA	719	
QY	5209	TGAAGGCTCTTTTCAGATGTAACTTATGAAGAAATATCTGCTTTG--TGTATATGCCA	5266	
DB	720	AGAAGCTCTTTTCA-AAGGAACCTTATGAAGAAATATCTCGCTTTGGCGCATTAATGCC	778	
QY	5267	GTTTGAATAC-TGGTCTTCAAAGTCTGTC-AAATTGTTATTTCAAGTGTG--CACAAAAACCA	5322	
DB	779	GCTAGATAACGGGTCTCTAAGTCGGGCAAAATGTTATTTCCGTGGGCACAAAAACCA	838	
QY	5323	GTTTTGAGGCTCTTAGACTTATAA--TTCCTTTGAAATAAACTGATAACTTATTTGTTATAA	5379	
DB	839	TTTTTGAAGGCTTAAACTAAAAATTTCTTTAAATAAAACTGAAAAACCTTATTTGTTAA	897	

RESULT 8
AL700151
LOCUS

DEFINITION DKFZp686I20117_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686I20117_5', mRNA sequence.
ACCESSION AL700151
VERSION AL700151.1 GI:19620684
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 740)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and
Wiemann,S.
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and
Wiemann,S.)
JOURNAL Unpublished (2001)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MedGenomix (Martinried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZp686I20117) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg; GERMANY; Email: clone@rzpd.de.
FEATURES
location/Qualifiers
1..740
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686I20117"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match	11.1%; Score 739; DB 9; Length 740;
Best Local Similarity	100.0%; Pred. No. 1.7e-136;
Matches 739; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	3742 TCTCATACCAAGTAATCCTTTCTTTGTGGAAATCTTATCCAGAAGCTTGGAGATCCCT 3801
Db	2 TCTCATACCAAGTAATCCTTTCTTTGTGGAAATCTTATCCAGAAGCTTGGAGATCCCT 61
QY	3802 GGCCAAAGTTTTTCCTATGCGTTTAATGTGAGATTGGTGCAGAAATAAAGCATATC 3861
Db	62 GGCCAAAGTTTTTCCTATGCGTTTAATGTGAGATTGGTGCAGAAATAAAGCATATC 121
QY	3862 CTGCTCCTCTAACCAAGCATCAGAGCGCGAAACCTCTTTTGGAGAAATAGGACACACTA 3921
Db	122 CTGCTCCTCTAACCAAGCATCAGAGCGCGAAACCTCTTTTGGAGAAATAGGACACACTA 181
QY	3922 TTATGAACCTACTTGTGTGACCTTCGAAATACCAAGTATACCTTGCATAATATAGATCAAC 3981
Db	182 TTATGAACCTACTTGTGTGACCTTCGAAATACCAAGTATACCTTGCATAATATAGATCAAC 241
QY	3982 TGTGTGATTCATATGGAAATGGGAAAAGCTGCATATAAAATACACGGAAAAGTACAGTG 4041
Db	242 TGTGTGATTCATATGGAAATGGGAAAAGCTGCATATAAAATACACGGAAAAGTACAGTG 301
QY	4042 ATGTAATGAAAGTACTTAAATTTCTTCCAATGAGCATGTCAATTAGCATTTGGACGAAGTTTCA 4101
Db	302 ATGTAATGAAAGTACTTAAATTTCTTCCAATGAGCATGTCAATTAGCATTTGGACGAAGTTTCA 361
QY	4102 GTACAGAACGAGATTCTCATCTAGCTGTGTATACAGAAATGATGGAAATTAATGAAAACAGG 4161
Db	362 GTACAGAACGAGATTCTCATCTAGCTGTGTATACAGAAATGATGGAAATTAATGAAAACAGG 421
QY	4162 CCAACAGTGCCACTGCCCATCCTAGAAAAGTACAGAGTCCAAGTTTTGTGGTATTCAAATG 4221

Db 422 CCAACAGTGCACCTGGCACTCTTAGAAGAGTGACAGTGCAGTTTGTGGTATTCATG 481
 QY 4222 GAGCTCTAAACAACTCTTTCAGGATTTCTGTAAAGTCCAGCATAGTTGAAGATGGCTTAA 4281
 Db 482 GAGCTCTAAACAACTCTTTCAGGATTTCTGTAAAGTCCAGCATAGTTGAAGATGGCTTAA 541
 QY 4282 TGGTACAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACGAAG 4341
 Db 542 TGGTACAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACGAAG 601
 QY 4342 ACTTTAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACGAATCT 4401
 Db 602 ACTTTAAATTAACCTCCAGAGACCATGATGGCTTCGGCTAGCTTTACGAGACACGAATCT 661
 QY 4402 CTGGGTAGATGCTCAAGAAAGAAACAAAGAGATTATCAGTTCAAGTGGATGGAATAT 4461
 Db 662 CTGGGTAGATGCTCAAGAAAGAAACAAAGAGATTATCAGTTCAAGTGGATGGAATAT 721
 QY 4462 CATTACAGGATTTCCAG 4480
 Db 722 CATTACAGGATTTCCAG 740

RESULT 9

BM976393/3
 LOCUS 723 bp MENA linear EST 21-FEB-2003
 DEFINITION UI-CF-EN1-ac2-h-01-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
 UI-CF-EN1-ac2-h-01-0-UI 3', mRNA sequence.
 ACCESSION BM976393
 VERSION 1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 723)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-23, >AT rich#Low complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source

1..723
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-ac2-h-01-0-UI"
 /tissue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="PH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-EN1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-EN1 is a normalized cDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dr primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CTGCTCAGGT.
 TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 6hr to LPS 24h
 TAG LIB=UI-CF-EN1
 TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 10.7%; Score 712; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1e-131;
 Matches 712; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4815 AGTTTACCATTAGAAATAGATTAGTGTCTTTCATTATAGAACATCTTTTGTAGTGAAG 4874
 Db 723 AGTTTACCATTAGAAATAGATTAGTGTCTTTCATTATAGAACATCTTTTGTAGTGAAG 664
 QY 4875 AATGTGCCATATTACATATTGCAACCTAAATTTGTTTAAACTTAACCTCCAGCATGAAGCTG 4934
 Db 663 AATGTGCCATATTACATATTGCAACCTAAATTTGTTTAAACTTAACCTCCAGCATGAAGCTG 604
 QY 4935 AAATGCCACAAACACATAAAGTATAAATATGCTCGATTTTGAACACATAGCTTTCCT 4994
 Db 603 AAATGCCACAAACACATAAAGTATAAATATGCTCGATTTTGAACACATAGCTTTCCT 544
 QY 4995 CTTTAGGAGGAGTAATGATCTTTTCAAAATCATTAGCACAATATTAAATATCTTAAATAATTA 5054
 Db 543 CTTTAGGAGGAGTAATGATCTTTTCAAAATCATTAGCACAATATTAAATATCTTAAATAATTA 484
 QY 5055 AGAGATCCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAGACAGAGATTTC 5114
 Db 483 AGAGATCCCATCTTTCTGTAGCTTTTACAAATTAATTTAAGTACTAAAGACAGAGATTTC 424
 QY 5115 TTTTAGAATTTATAGCATTTTACTGTGTATTATTAATGCTAAGCCAAAGTATCTGCAC 5174
 Db 423 TTTTAGAATTTATAGCATTTTACTGTGTATTATTAATGCTAAGCCAAAGTATCTGCAC 364
 QY 5175 TAGGTATACCTCTTTATGCCAATAATGATTTTAAAGAGGCTCTTTTTCAGATGTAACCTT 5234
 Db 363 TAGGTATACCTCTTTATGCCAATAATGATTTTAAAGAGGCTCTTTTTCAGATGTAACCTT 304
 QY 5235 ATCAAGAAATATCTGCTTTGTATATGCCAGTATAGATACTGTTTCTTAAAGTCTGTC 5294
 Db 303 ATCAAGAAATATCTGCTTTGTATATGCCAGTATAGATACTGTTTCTTAAAGTCTGTC 244
 QY 5295 AAATTCGTATTTTCAGTGGCACAAAAACAGTTTGTAGGCTTTAGACTTATAATTCCTTGA 5354
 Db 243 AAATTCGTATTTTCAGTGGCACAAAAACAGTTTGTAGGCTTTAGACTTATAATTCCTTGA 184
 QY 5355 TAAACCTGATACTTTATTTGTATTAATTTGGAGTGGAGACCTCTCCATAATTAGATAAAC 5414
 Db 183 TAAACCTGATACTTTATTTGTATTAATTTGGAGTGGAGACCTCTCCATAATTAGATAAAC 124
 QY 5415 TCCTTTTGGATTATAATCAGAAATTTTGCCTTTTCTTCTCAAAATTTATACATATGATG 5474
 Db 123 TCCTTTTGGATTATAATCAGAAATTTTGCCTTTTCTTCTCAAAATTTATACATATGATG 64
 QY 5475 TATTATATATCCACATATATAGTTTTCCTCTGATTAATATCGATATTAATAA 5526
 Db 63 TATTATATATCCACATATATAGTTTTCCTCTGATTAATATCGATATTAATAA 12

RESULT 10

```

AUI33002
LOCUS       AUI33002 NT2RP4 Homo sapiens cDNA clone NT2RP4001068 5', mRNA          linear          EST 01-AUG-2002
DEFINITION   AUI33002 NT2RP4 Homo sapiens cDNA clone NT2RP4001068 5', mRNA
ACCESSION   AUI33002
VERSION     AUI33002.1 GI:10993541
SOURCE      HRI human cDNA project
ORGANISM    Homo sapiens (human)
REFERENCE   1 (bases 1 to 727)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y., Saito, K., Yamamoto, J.,
            Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuho, Y. and
            Isogai, T.
            HRI human cDNA project (Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y.,
            Saito, K., Yamamoto, J., Nishikawa, T., Nakamura, Y., Nagai, T.,
            Sugano, S., Masuho, Y., Isogai, T.)
            Unpublished (2000)
JOURNAL     Contact: Takao Isogai
COMMENT     Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES    Location/Qualifiers
             1..727
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /c1one="NT2RP4001068"
             /cell_type="teratocarcinoma"
             /cell_line="NT2"
             /clone_lib="NT2RP4"
             /note="vector: pME18SFL3; mRNA from NT2 neuronal precursor
             cells after 2-weeks retinoic acid (RA) induction"
ORIGIN
Query Match          10.7%; Score 708.8; DB 9; Length 727;
Best Local Similarity 99.0%; Pred. No. 1.8e-130;
Matches 721; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 4965  GTCTGATTTTGAACACATAAGCTTTTGCTTTTAGCGAGGAATGATCTTTTCAAAATCAT 5024
DB 1  GTCTGATTTTGAACACATAAGCTTTTGCTTTTAGCGAGGAATGATCTTTTCAAAATCAT 60
QY 5025  TAGCACATATTTAAATATCTMAAAATTTAAGAGATCCATCTTCTGTAGCTTTACAAT 5084
DB 61  TAGCACATATTTAAATATCTMAAAATTTAAGAGATCCATCTTCTGTAGCTTTACAAT 120
QY 5085  TAATTTAAGTACTAAAAGACAAGGATTTCTTTTAAGAAATTTATAGCAATTTACTGTGTT 5144
DB 121  TAATTTAAGTACTAAAAGACAAGGATTTCTTTTAAGAAATTTATAGCAATTTACTGTGTT 180
QY 5145  ATTTAAATGCTAACCCAAAGTATCTGCACCTAGGTATACCTTTTATGCCAATAATGATT 5204
DB 181  ATTTAAATGCTAACCCAAAGTATCTGCACCTAGGTATACCTTTTATGCCAATAATGATT 240
QY 5205  TTAATGAAGGCTCTTTTCAGATGTAACCTTATGAAGGAATATCTGCTTTGTGTATATGC 5264
DB 241  TTAATGAAGGCTCTTTTCAGATGTAACCTTATGAAGGAATATCTGCTTTGTGTATATGC 300
QY 5265  CAGTTAGAACTACTGGTTTCTAAAGTCTGTCAAAATGTTATTCATGTGCGACAAAACCAAGT 5324
DB 301  CAGTTAGAACTACTGGTTTCTAAAGTCTGTCAAAATGTTATTCATGTGCGACAAAACCAAGT 360
QY 5325  TTTCAGGCTCTAGACTTATATCTTTTGAATATAAAGTAACTTATTCGTATTAATTGGA 5384

```

```

DB 361  TTTGAGGCTTTAGACTTATAATTTCTTTGAATAAAACGTATAACTTATTTTGTATAATTGGA 420
QY 5385  GTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATATAATCAGAAATTTTGCT 5444
DB 421  GTGGAGACCTACCTCCATAATTAGATAAACTCTTTTGGATTATATAATCAGAAATTTTGCT 480
QY 5445  TTTTCTTCTCAAATTTATACATATGTATGTATATATATATATATCCACATATATAGTTTCCCT 5504
DB 481  TTTTCTTCTCAAATTTATACATATGTATGTATATATATATATATCCACATATATAGTTTCCCT 540
QY 5505  GATTAAATCGATTAATAAATATTCGGGTGCTTCAGGACTTTTGTCTTCTATATTTAAG 5564
DB 541  GATTAAATCGATTAATAAATATTCGGGTGCTTCAGGACTTTTGTCTTCTATATTTAAG 600
QY 5565  TATATTGTTTTATAGCAAGACATATTTCTGAATGTTTTATAAACTTTTAAATAATTTATA 5624
DB 601  TATATTGTTTTATAGCAAGACATATTTCTGAATGTTTTATAAACTTTTAAATAATTTATA 660
QY 5625  TGTAGTAATATTTTGTATCAATGATATTTTTCCTCTCTTCCTTCCTTCCTTCCTTCCTTCCT 5684
DB 661  TGTAGTAATATTTTGTATCAATGATATTTTTCCTCTCTTCCTTCCTTCCTTCCTTCCTTCCT 719
QY 5685  TACCACCTG 5692
DB 720  TACCACCTG 727

RESULT 11
LOCUS    AL596824
DEFINITION   717 bp mRNA linear EST 04-SEP-2003
            DKF2p451N191.r1.451 (synonym: hlcc1) spinal cord Homo sapiens cDNA
            clone DKF2p451N191.5', mRNA sequence.
ACCESSION   AL596824
VERSION     AL596824.1 GI:15154520
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Poustka, A., Wellenreuther, R., Mewes, H.W., Weil, B. and Wiemann, S.).
            EST (Poustka, A., Wellenreuther, R., Mewes, H.W., Weil, B. and
            Wiemann, S.)
            Unpublished (1999)
            Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No si sequence available.
            This clone (DKF2p451N191) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES    Location/Qualifiers
             1..717
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="DKF2p451N191"
             /tissue_type="human spinal cord"
             /dev_stage="adult"
             /lab_host="DH10B"
             /clone_lib="451 (synonym: hlcc1) spinal cord"
             /note="vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN
Query Match          10.4%; Score 687.2; DB 9; Length 717;
Best Local Similarity 99.6%; Pred. No. 3.5e-126;
Matches 689; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

628 CTGATCTGATAGTGAACATGGTAACTTGTCTATGCAACCAATAGTGAAGAGATTTA 687
 Db 26 GGGATCTGATAGTGAACATGGTAACTTGTCTATGCAACCAATAGTGAAGAGATTTA 85
 Qy 688 AAAAATTTATCCAGATGATTTTAACTGCTAATGAGATTCCTGATGGATTTGATTTAT 747
 Db 86 AAAAATTTATCCAGATGATTTTAACTGCTAATGAGATTCCTGATGGATTTGATTTAT 145
 Qy 748 CTTGAGTCTGATAGTCCCTGCTGTTCTTCAACAGACCATGATGATGATCTGTCAGAG 807
 Db 146 CTTGAGTCTGATAGTCCCTGCTGTTCTTCAACAGACCATGATGATGATCTGTCAGAG 205
 Qy 808 AACACAGAAATGATACATGCTCTCAATACAAATACAAATAGGAATCGGAGGAATCAAGAAT 867
 Db 206 AACACAGAAATGATACATGCTCTCAATACAAATAGGAATCGGAGGAATCAAGAAT 265
 Qy 868 TGGGTATATAAGTAGATACACACTTTTCAAGATTCCTATTAATTAAGTGGACAGAAAT 927
 Db 266 TGGGTATATAAGTAGATACACACTTTTCAAGATTCCTATTAATTAAGTGGACAGAAAT 325
 Qy 928 TAAAGATATAAGATCTTTTAACTGCTAATGAGATTCCTGATGGATTTTAACTGTCATCTG 987
 Db 326 TAAAGATATAAGATCTTTTAACTGCTAATGAGATTCCTGATGGATTTTAACTGTCATCTG 385
 Qy 988 CTTTGAATCGAACAAGTTTCCAAATGTTTCAATGCAAGACCAAGCTACACACAGAGCC 1047
 Db 386 CTTTGAATCGAACAAGTTTCCAAATGTTTCAATGCAAGACCAAGCTACACACAGAGCC 445
 Qy 1048 AGCCATGCTGATTAATAAGATGTTTCAAGATTCCTATTAATTAAGTGGACAGTCA 1107
 Db 446 AGCCATGCTGATTAATAAGATGTTTCAAGATTCCTATTAATTAAGTGGACAGTCA 505
 Qy 1108 TAACTGCGCGAAGTGTAAAGAGAGGCGCAAGAGCTTTTCACTGACCTGACGCTTC 1167
 Db 506 TAACTGCGCGAAGTGTAAAGAGAGGCGCAAGAGCTTTTCACTGACCTGACGCTTC 565
 Qy 1168 CGAAATTAAGATTTATGCTTAAATGATTCAAATTCAGAGATGAAATTTCAAAATTA 1227
 Db 566 CGAAATTAAGATTTATGCTTAAATGATTCAAATTCAGAGATGAAATTTCAAAATTA 625
 Qy 1228 CTGATTTTCTTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
 Db 626 CTGATTTTCTTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 685
 Qy 1288 AAAGTTTAGACCTTAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1319
 Db 686 AAAGTTTAGACCTTAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 717

RESULT 12
 CD557100
 LOCUS
 DEFINITION
 AGENCOURT 14413152 NIH_MGC_180 Homo sapiens cdna clone
 IMAGE:30390481 5', mRNA sequence.
 CD557100
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 EST.
 CD557100.1 GI:31593168
 Homo sapiens (human)
 ORGANISM

935 bp mRNA linear EST 11-JUN-2003
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 935)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhardt, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NTH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cdna Library Preparation: Invitrogen Corp

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM465 row: k column: 02
 High quality sequence start: 24
 High quality sequence stop: 620.
 Location/Qualifiers
 1..935
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30390481"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC_180"
 /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site.1: NotI;
 Site.2: EcoRV (destroyed); Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.68 kb. Library was
 constructed by (Invitrogen). Note: this is a NIH_MGC
 Library."

ORIGIN

Query Match 10.3%; Score 682.6; DB 14; Length 935;
 Best Local Similarity 98.7%; Pred. No. 2.8e-125;
 Matches 741; Conservative 0; Mismatches 4; Indels 6; Gaps 5;
 Qy 3303 GAGGACAGTTCCGCCCTCTGCTGTCATCTCGAGAAAGAGGATCAGTCTGTAGTA 3362
 Db 33 GATGACAGTTTCCGCCCTCTGCTGTCATCTCGAGAAAGAGGATCAGTCTGTAGTA 92
 Qy 3363 GAAGAACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGTGAAGGCTTCATCCT 3422
 Db 93 GAAGAACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGTGAAGGCTTCATCCT 152
 Qy 3423 GTTACATTTGCTTAATGCTTAATCTCTCGTGAATGTCAAATTCATATTTTATTCCTCA 3482
 Db 153 GTTACATTTGCTTAATGCTTAATCTCTCGTGAATGTCAAATTCATATTTTATTCCTCA 212
 Qy 3483 GACAAATATTTGCTTCTTCAACCAATGATTCGCTTGGGACAGGCAAAATTTAT 3542
 Db 213 GACAAATATTTGCTTCTTCAACCAATGATTCGCTTGGGACAGGCAAAATTTAT 272
 Qy 3543 ATCTATTTGTTAGTTTGGCAATGAGATCTATCTTCTTGAAGTGAAGGCTTCATCCT 3602
 Db 273 ATCTATTTGTTAGTTTGGCAATGAGATCTATCTTCTTGAAGTGAAGGCTTCATCCT 332
 Qy 3603 ATCCACCATATATAAGGATGCTCTTAAAGGAAATACATAGAAACTTGGCAATATTACC 3662
 Db 333 ATCCACCATATATAAGGATGCTCTTAAAGGAAATACATAGAAACTTGGCAATATTACC 392
 Qy 3663 TTCTAGAGAGTTTCTCAGTAGAGAGATCAGGAGGATTCCTGTTTATTCACCTACT 3722
 Db 393 TTCTAGAGAGTTTCTCAGTAGAGAGATCAGGAGGATTCCTGTTTATTCACCTACT 452
 Qy 3723 TTCTAGAAACTTGATGATCTCTCATTACCAAGTAACTCTTTTCTTGTGGAAATCTTATC 3782
 Db 453 TTCTAGAAACTTGATGATCTCTCATTACCAAGTAACTCTTTTCTTGTGGAAATCTTATC 512
 Qy 3783 CAGAAGCTTTGAGATTCCTTGGGCAAGGTTTTCCTATCGGTTTAAATGTTGAGATTTGGT 3842
 Db 513 CAGAAGCTTTGAGATTCCTTGGGCAAGGTTTTCCTATCGGTTTAAATGTTGAGATTTGGT 572
 Qy 3843 CGAGATATAAGAGCATATCTCTCTCTTCAACAGATCAG-AGGCCGAAACCTCTTTT 3901
 Db 573 CGAGATATAAGAGCATATCTCTCTCTTCAACAGATCAGAGGCCGAAACCTCTTTT 632
 Qy 3902 TGGAGAAATAGGACACACATTTATGAACCTTACTTG- TTGACCTTCGAAATACCAAGTATA 3960
 Db 633 TGGAGAAATAGGACACACATTTATGAACCTTACTTGTTTGACCTTCGAAATACCAAGTATA 692
 Qy 3961 CCTTGCATATATAGATCAACT-GTTGATTCATATG- -AATATGGGAAAGCTGCATAA 4017

Db	693	CTTGTATATATAGTCACTGGTGGTATCATATGGAAATGGGAAAGAGTGCATTA	752
Qy	4018	AAATACCACGG-AAAAAGTACAGTGATGTA	4047
Db	753	AAATACCACGGAAAAAGTACAGTGATGTA	783
RESULT 13			
CB306662/c			
LOCUS	713 bp	mRNA	linear
DEFINITION	UI-CF-FNO-aem-d-09-0-UI-s1 UI-CF-FNO Homo sapiens cDNA clone		EST 04-MAR-2003
ACCESSION	CB306662		
VERSION	CB306662.1	GI:28847173	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 713)		
AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.		
TITLE	Normalization and subtraction: two approaches to facilitate gene		
JOURNAL	discovery		
MEDLINE	Genome Res. 6 (9), 791-806 (1996)		
PUBMED	97044477		
COMMENT	8889548		
	Contact: McCray, PB		
	McCray Lab		
	University of Iowa		
	2024 University of Iowa Med Labs, Iowa City, IA 52242, USA		
	Tel: 319 356 4866		
	Fax: 319 356 7171		
	Email: paul-mccray@uiowa.edu		
	Tissue Procurement: Dr. M. J. Welsh, University of Iowa		
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: Researchers may obtain clones from Research		
	Genetics (www.resgen.com) or from Open Biosystems		
	(www.openbiosystems.com).		
	Seq primer: M13 FORWARD		
	POLYA=Yes.		
FEATURES	Location/Qualifiers		
source	1. .713		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="UI-CF-FNO-aem-d-09-0-UI"		
	/tissue_type="Human Lung Epithelial cells"		
	/lab_host="DH10B (Life Technologies) (T1 phage resistant)"		
	/clone_lib="UI-CF-FNO"		
	/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a		
	modified polylinker; Site 1: EcoR I; Site 2: Not I;		
	UI-CF-FNO is a subtracted cDNA library derived from two		
	normalized Human lung epithelial cell libraries (EN1 and		
	DUI1) The library was subtracted according to		
	Bonaldo, Lennon and Soares, Genome Research, 6:791-806,		
	1996. For additional information, contact:		
	bento-soares@uiowa.edu		
	TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368		
	TAG LIB=UI-CF-FNO		
	TAG_SEQ=GGCTGTAGGC		
ORIGIN			
Query Match	10.2%	Score 679.4;	DB 14; Length 713;
Best Local Similarity	99.7%	Pred. No. 1.2e-124;	
Matches	680;	Conservative 0;	Mismatches 2; Indels 0; Gaps 0;
Qy	5949	TGATATTCATTTGTACAGAGCTTCTCTTACTTCACACACGCAAAAAGTGGGGGC	6008
Db	698	TGATATTCATTTGTACAGAGCTTCTCTTACTTCACACACGCAAAAAGTGGGGGC	639

RESULT 14
AW953665/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AW953665
EST365735
AW953665
AW953665.1
EST.
Homo sapiens (human)
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 711)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C.,
Holt, J.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 53
Seq primer: Reverse.
Location/Qualifiers
1. .711

ORIGIN

Query Match 10.2%; Score 676.4; DB 10; Length 711;
 Best Local Similarity 97.7%; Pred. No. 4.9e-124;
 Matches 586; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE resequences, VAGC"
 /note="Vector: pBluescriptSKm"

QY 4842 TTTTTCATTATAGAACATCTTTTATAGTGAAGAATGGCCATATACATATGCAACCT 4901
 DB 702 TTTTCCATAAAAAAATCTTTTATAGTGAAGAATGGCCATATATCATATGCAACCT 643

QY 4902 AATTGTTTAAACTAATCCAGCACTAAAGCTGAAATGCCACAACACTAAAGTATAAA 4961
 DB 642 ATTTTGTTTAAACTAATCCAGCACTAAAGTGAATGCCCTCAACACTAAAGTATAAA 583

QY 4962 TATGTCGATTTTGAACACATAAGCTTTGCTCTTTAGGCGAGGAATGATCTTTTCAAA 5021
 DB 582 TATGTCGATTTTGAACACATAAGCTTTGCTCTTTAGGCGAGGAATGATCTTTTCAAA 523

QY 5022 CATTTAGCAATATTTAATATCTTAAATTTAAGATCCATCTTTCTGAGCTTTTAC 5081
 DB 522 CATTTAGCAATATTTAATATCTTAAATTTAAGATCCATCTTTCTGAGCTTTTAC 463

QY 5082 AATTAAATTTAAGTACTAAAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGT 5141
 DB 462 AATTAAATTTAAGTCTTAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGT 403

QY 5142 GTTATTTAAATGCTTAAGCAAGTAATCTGCACTTAGGTATACCTTTTATGCCAATTAATG 5201
 DB 402 GTTATTTAAATGCTTAAGCAAGTAATCTGCACTTAGGTATACCTTTTATGCCAATTAATG 343

QY 5202 ATTTTAAATGAAGGCTTTTTCAGATGTAACCTTTATGAAGAAATATCTGCTTTCTGTATA 5261
 DB 342 ATTTTAAATGAAGGCTTTTTCAGATGTAACCTTTATGAAGAAATATCTGCTTTCTGTATA 283

QY 5262 TGCCAGTTTGAATACTGTTTCTTAAAGTCTGCAAAATGTTATTTTCAAGTGCACAAAAACC 5321
 DB 282 TGCCAGTTTGAATACTGTTTCTTAAAGTCTGCAAAATGTTATTTTCAAGTGCACAAAAACC 223

QY 5322 AGTTTGAAGTCTTAGACTTATTAATCTTGAATTAAGCACTGATTAATTTTGTATTAAT 5381
 DB 222 AGTTTGAAGTCTTAGACTTATTAATCTTGAATTAAGCACTGATTAATTTTGTATTAAT 163

QY 5382 GGAGTGGAGACTACCTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTG 5441
 DB 162 GGAGTGGAGACTACCTCCATAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTG 103

QY 5442 CTTTTTCTTCTGAAATTTACATATGATGATGATATATATATATCCATATATAGTTTTC 5501
 DB 102 CTTTTTCTTCTGAAATTTACATATGATGATGATATATATATATCCATATATAGTTTTC 43

QY 5502 CTTGATTAATGGATATTAATAATTAATGCGGGTCTCTCAGGA 5543
 DB 42 CTTGATTAATGGATATTAATAATTAATGCGGGTCTCTCAGGA 1

RESULT 15
 CB962678/c
 LOCUS
 DEFINITION AGNCOURT 13465015 NIH MGC 187 Homo sapiens cDNA clone
 IMAGE:30318937 5', mRNA sequence.
 ACCESSION CB962678
 VERSION CB962678.1 GI:30218796
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 758)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapps@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDCM114 row: f column: 02
 High quality sequence stop: 611.

FEATURES

1. .758
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30318937"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH MGC 187"
 /note="Organ: Blood_vessels - aorta, basilar and artery;
 Vector: pDNR-LIB; Site 1: SfiI (ggccattatggc); Site 2:
 SfiI (ggccgctcggcc); 5' and 3' adaptors were used in
 cloning as follows: 5' adaptor sequence:
 5'-CACGCGCATATGCGC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGCGCGCACATG-DT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.4 kb
 (range 0.5-4.0 kb). 14/15 colonies contained inserts by
 PCR. This library was enriched for full-length clones and
 was constructed by Clontech Laboratories (Palo Alto, CA).
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.1%; Score 668.4; DB 14; Length 758;
 Best Local Similarity 99.7%; Pred. No. 1.9e-122;
 Matches 669; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 679 AAGATATTAATAAATATTGCGAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATGAT 738
 DB 672 AAGATATTAATAAATATTGCGAGATGATTTTAAGTCTTAATGCGAGATTCCTTGATGAT 613

QY 739 TGGATTTATCTTCAGTGTGAGATCTCCCTGTGTTCTTCAACGACCATGATAGTATA 798
 DB 612 TGGATTTATCTTCAGTGTGAGATCTCCCTGTGTTCTTCAACGACCATGATAGTATA 553

QY 799 CTGTCAGAGAACACAGAAATGATATCAGTCTTGAATTAACAAATAGAGAAATCGAGGAA 858
 DB 552 CTGTCAGAGAACACAGAAATGATATCAGTCTTGAATTAACAAATAGAGAAATCGAGGAA 493

QY 859 TCAAGAAATGGGTAAAGTAGATACACACTTTCAGATTCCTATTAATACAGTGA 918
 DB 492 TCAAGAAATGGGTAAAGTAGATACACACTTTCAGATTCCTATTAATACAGTGA 433

QY 919 CAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAATCAATCTTGATTTTACA 978
 DB 432 CAGAAATTTAAAGATATAAAGATCTTTAATCAGTTAGAATCAATCTTGATTTTACA 373

QY 979 TGTCTCTGCTTTGACTGACAAAGTTCAAAATGTTTCNTGCCAAGACAGCTACAC 1038
 DB 372 TGTCTCTGCTTTGACTGACAAAGTTCAAAATGTTTCNTGCCAAGACAGCTACAC 313

QY 1039 ACAAGAGCCAGCATCTGGATTTACTTAAAGATGTTGGCTTAGTAAAGAGGAAGTAGATG 1098
 DB 312 ACAAGAGCCAGCATCTGGATTTACTTAAAGATGTTGGCTTAGTAAAGAGGAAGTAGATG 253

QY 1099 TGGAGTGCATAACTCCGCGAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTTGACCT 1158
 DB 252 TGGAGTGCATAACTCCGCGAGAAATGTTTAAAGAGAGGCGCAAGCAAGTCTTTGACCT 193

QY 1159 CGAGCTCTCCGAAAATGAAGATTTATCTTAAATGATTTCAAAATTCAGAGATGAAAT 1218

```
Db      192 GCAGCCTTCGAAAAATGAAGATTATGCTTAAATGATTCAAATTCAGAGATGAAAAATT 133
Qy      1219 TCAAAATFACCTGACTTTTCCTTTCAGGAGAGATAAGACTGTATATAAAACAATCTGCACAAG 1278
        |||||
Db      132  TCAAAATFACCTGACTTTTCCTTTCAGGAGAGATAAGACTGTATATAAAACAATCTGCACAAG 73
        |||||
Qy      1279 AAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCCCAAGATTCTCTTCAGCTT 1338
        |||||
Db      72  AAGACTCAAAAAGTTTAGACCTTAAGGATAATGATGTAATCCCAAGATTCTCTTCAGCTT 13
        |||||
Qy      1339 TACATGTTTCC 1349
        |||||
Db      12  TACATGTTTCC 2
```

Search completed: May 18, 2004, 07:03:30
Job time : 10173.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 16:37:02 ; Search time 290.233 Seconds
(without alignments)
12680.960 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 6632

Sequence: 1 actccggcggggtagctc.....ttggaggcacatttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgm2_6/ptodata/2/ina/PTCUS_COMB.seq.*
- 6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.4	1.0	13124	2	US-08-487-826B-13
2	62.8	0.9	8961	4	US-10-204-708-80
3	61	0.9	6152	3	US-08-973-462-1
4	59	0.9	19124	2	US-08-487-826B-13
5	57.8	0.9	5666	4	US-10-204-708-30
6	57.4	0.9	6156	4	US-10-204-708-59
7	56.8	0.9	7218	1	US-08-232-433-14
8	56.8	0.9	53332	4	US-09-801-881-3
9	56.4	0.9	640681	4	US-09-790-988-1
10	56	0.8	6124	4	US-08-213-419B-3
11	55.2	0.8	636	3	US-08-998-416-1137
12	55.2	0.8	837	3	US-08-998-416-288
13	53.8	0.8	6113	4	US-10-204-708-14
14	53.6	0.8	6156	4	US-10-204-708-60
15	53	0.8	832	4	US-09-521-976-2813
16	52.8	0.8	640681	4	US-09-790-988-1
17	52.6	0.8	665	2	US-08-883-795A-36
18	52	0.8	615	3	US-08-998-416-186
19	52	0.8	6866	4	US-10-204-708-19
20	51.6	0.8	2095	6	5231168-1
21	51.4	0.8	5852	1	US-07-867-106-2
22	51.2	0.8	658	3	US-08-998-416-595
23	51.2	0.8	5340	4	US-09-627-122-21
24	51	0.8	20674	4	US-09-641-638-651
25	50.4	0.8	832	4	US-09-621-976-2813
26	50.4	0.8	6243	2	US-09-056-075-1
27	50.2	0.8	6801	4	US-10-204-708-61

28	49.6	0.7	6113	4	US-10-204-708-13	Sequence 13, Appl
29	49.4	0.7	11049	4	US-10-204-708-22	Sequence 22, Appl
30	49	0.7	1183	2	US-08-731-722-8	Sequence 8, Appl
31	49	0.7	4140	3	US-08-894-731-2	Sequence 2, Appl
32	49	0.7	6317	4	US-10-204-708-11	Sequence 11, Appl
33	48.8	0.7	10467	4	US-10-204-708-2	Sequence 2, Appl
34	48.6	0.7	5152	4	US-10-204-708-73	Sequence 73, Appl
35	48.6	0.7	168575	4	US-09-426-290-1	Sequence 1, Appl
36	48.4	0.7	724	3	US-08-998-416-683	Sequence 683, Appl
37	48.4	0.7	6070	4	US-10-204-708-10	Sequence 10, Appl
38	48.4	0.7	6326	4	US-10-204-708-57	Sequence 57, Appl
39	48.2	0.7	767	3	US-08-998-416-472	Sequence 472, Appl
40	48.2	0.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
41	48.2	0.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
42	48.2	0.7	1511	2	US-08-544-332-8	Sequence 8, Appl
43	48.2	0.7	1511	4	US-09-370-861A-8	Sequence 8, Appl
44	48.2	0.7	6768	1	US-08-107-755A-1	Sequence 1, Appl
45	48.2	0.7	8457	1	US-07-991-867B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-487-826B-13
; Sequence 13, Application US/08487826B

; Patent No. 593827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellens, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israel, Ned

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH121.001CPI

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 19124 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; US-08-487-826B-13

Query Match 1.0%; Score 64.4; DB 2; Length 19124;

Best Local Similarity 45.3%; Pred. No. 6.4e-05;

QY 5790 TCACATTTAATCTAGAAATCTTGGCACTTCGCATATTTTCATTCGACACTGTGTGA 5849
Db 6103 TTTTTCCTTTTCGTTTATTTGTAATATTTAAATTTTTCGTTTAAAGTATAT 6162
QY 5850 TAAGTATAAATTTAAAGAACTAATTTACTTTTCATATTTTAAATTTCTTTATAT 5903
Db 6163 TTGTAATAAAGTTTTCGATTTTGGGATTTAGTTTTCATATTTTATATTTATTT 6216

RESULT 3
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRULHE, PIERRE
; APPLICANT: DUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match 0.9%; Score 61; DB 3; Length 6152;
Best Local Similarity 48.7%; Pred. No. 0.00025;
Matches 166; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
QY 5336 AGACTTAATCTTTGAAATAAACTGATTAACCTTTATTTGATATTTGAGTGGAGACCTA 5395
Db 5716 AT 5775
QY 5396 CCTCCATAATPAGATAAACTCTTTTGGATATATATATATATATATATATATATAT 5455
Db 5776 TATTAATATTTTATATTTTCCATATATATATATATATATATATATATATATATAT 5835
QY 5456 AAATTTATACAT 5515
Db 5836 AATGTTTTTACAGATTTATGTTTTTAAATTAATATATATATATATATATATATAT 5895
QY 5516 TATTAATAATATTCGGGTGCTTCAGGACTTTTTCCTTCTATATTTTAAAGTATATGTTT 5575
Db 5896 TATTTATTCAGAT 5955
QY 5576 TATAGCAAGACATATCTGAAATGTTTTTATATATATATATATATATATATATATAT 5635
Db 5956 TAAT 6015
QY 5636 TTTTGTATCAATGCAATTTATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5676
Db 6016 TTGTATCAATAAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 6056

RESULT 4
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
Query Match 0.9%; Score 59; DB 2; Length 19124;
Best Local Similarity 43.0%; Pred. No. 0.0011;
Matches 504; Conservative 0; Mismatches 660; Indels 9; Gaps 4;
QY 5035 TTTAAATATCTAAATTTAAGAGATCCATATCTTCTGTAGCTTTACATTAATTTAAGT 5094
Db 7062 TATTATATATACAT 7003
QY 5095 ACTAAAACACAGAGATTTCTTTAAGAAATTTATAGCAATTTTACTGTGTTTAAATGC 5154
Db 7002 ATATATTTAGTTTAT 6943
QY 5155 TAAGCCAAAGTATCTGCACCTTAGGTATACCTCTTTATGCCAATAATGATTTTAAAGAG 5214
Db 6942 TAAATTTAAATATCTTACAAAAAATAATAATCAGAAATATATATATTTATGTATTC 6883
QY 5215 CTCTTTTCAGATGAACCTTTATCAAGGAATATCTCTTTGTGTATATGCCAGTTAGAAT 5274
Db 6882 CTTTATTTATCTATTTAATTTATATATATATATATATATATATATATATATATAT 5823
QY 5275 ACTGGTTTCTAAAGTCTGTCAAAATTTGTTTTCAGTGGCAAAAAACAGTTTTCAGGTCT 5334
Db 5822 AATTAATAATGAAAAAATAATAATCGAAAAATACAAACATATAAAAAAGTATATATGCAA 6763
QY 5335 TAGACTTTAATTTCTTTGAATAAACTGATACTTTATTTGTATATTTGAGTGGAG---A 5391
Db 6762 CGTGTATATATATTTAATTTAATTAACATTAATATATATATATATTTTTCACATTTATTTA 6703
QY 5392 CCTACCTCCATATATTAGATAAATCTTTTGGATTATATATCAGAAATTTTGCCTTTTCTCT 5451
Db 6702 ATTAT 6643
QY 5452 TCTCAAAATTTATCAT 5511
Db 6642 TCCTATCTATATTTATCTATCATATTTATATATATATATATATATATATATATATATAT 6583

5512 TCGATATTAAATAATTCGGGTGCTTCAGGACTTTTTCCTCTATATTTAAAGTATATTG 5571
Db ATATCTTTTGTATTTGTTATTAATAAGTAGATATATATATTTATTTTCTT 6523
5572 TTTTATACCAAGAACATATTCGAATGTTTATATAATCTTTAATAATTATATGAGGT 5631
Db GTTATATATTAAATTTATATATAATAATATATAGTACCAAAAAAATGATGTAAT 6463
5632 ATATTTTGTATCAATGCAATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5691
Db AAATAACAGGAATAAT 6403
5692 GTATTACCACTTCAAGAGTACGACGAGCGGCAGATGACCTTGAAGTAGTCATTA 5751
Db TCAATTTGT--TTTGTGTTATTTTATGTTTTCATGCAATTTATGAATTTCAAATTT 6345
5752 TGTAGCAATAATGAAGCTGAAACAGGTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 5811
Db TATTGTATATAATAAAAAAATAAGTAAAAATACACATTTATAATATATATATATATAT 6285
5812 TCTTGGCAACTTCGCATATTTTCATTCAGCTGATGATATAGTATAATTTAAATGAAT 5871
Db GAGTTATTAATAAATGTTCTCTATATATTTATATAAAGTAAATATTTGTTA--- 6228
5872 AATTAATCTTTGATATTTAAATCTTTATATGAGTATTTTCTCTCTCTCTCTCTCTCTCT 5931
Db TAAATATAATACATATATCTCTATATAATAATATTAATATCTTTAAAGTATATACTA 6168
5932 ACATAGTTAAATCTGATTTGAAATCTGACAGCTTTCAGAGCTTCTCTCTCTCTCTCTCT 5991
Db AAATATAATAAATGCAATGTA-TAAAAATAGTATAAATCAATCATATATATATATATAT 6109
5992 GCAAAAAAGTGGGGGCATATTTAGTCTCTGTCATTTAAGTTATGTAATAAATTTAAATCA 6051
Db AT 6049
6052 TTATTTTCTGATCTTTAAACATCTCATGTGTAATATATATATATATATATATATATAT 6111
Db TATCCCAATCTGATATGTTGTTTATATTTGTTATATATATATATATATATATATATATAT 5999
6112 TATATTTCAAGAAAAAGAAATATGTTAAATAGCCCTGTTTAAAGAAAAATATTTATGAA 6171
Db AGAGTAAACAAAAAAT 5988
6172 GCATCTCACTTGAAGTCAAGTCAAGTTATA 6204
Db TTAGTATAGTTTATGTTCTTTATATATATATA 5896

RESULT 5
US-10-204-708-30
; Sequence 30, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE OF INVENTION: By Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10204708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98

; SEQ ID NO 30
; LENGTH: 5666
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-30

Query Match 0.9%; Score 57.8; DB 4; Length 5666;
Best Local Similarity 45.7%; Pred. No. 0.0013;
Matches 293; Conservative 0; Mismatches 332; Indels 16; Gaps 2;

QY 5067 TTTCTCTAGCTTTACAAATTAATTTAAAGTACTAAAAAGACAGGATTTCTTTTAAAGAAAT 5126
Db TTTGGAAATATTAAGATTTGTTAGGAGTTTGTAGGTTAGAAATTTTATATAATAT 3367
QY 5127 TATAGCAATTTACTGCTTATTTTAAATGCTAAGCAAGTATCTGCACCTAGGTATACCTC 5186
Db TTAATAATTTGATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3427
QY 5187 TTTATGCCAATATGATTTTAAATGAAGCTCTTTTTCAGATGTAACCTTATGAAGAAAT- 5245
Db ATTATATGATATATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3487
QY 5246 -----ATCTGCTTTGTATATGCCAGTTAGAATACCTGTTTCTTAAAGTCTGCA 5295
Db GTTTTATTAAGTTGATAGTAGTGAAGCAATTTGTAATAATTTAGTAATTTATTTTATTTT 3547
QY 5296 AATTTGATTTTCAGTGGCACAACCAACCGTTTGGAGTCTTAGACCTATATAATCTTTGAAT 5355
Db TATTTTCTTTGTTTGTAAAGATATAGTTATGTTTATATATATATATATATATATATATAT 3607
QY 5356 AAAAAGTGAATCTTATTTGTTATATTTGAGTGGAGACCTACCTCCATTAATTTAGATAAACT 5415
Db ATAACTGGTTTAAATTTGTTATTTTAAAT-----AAATTAATAGTAAATAGGTTAAAT 3662
QY 5416 CTTTITGGATTAATATCAAGATTTTGCCTTTTCTCTCAATTTATACATATGTTGT 5475
Db ATATTAGTTTAAATTTTAAATATAAATAATATATAGTTAAATTTATATATATATATAT 3722
QY 5476 ATTATATATCCATATATATAGTTTCCCTGATTAATGAATATTAATAATTTAAATTTGGGGTG 5535
Db TTTTGAAGTTTCTAGTGAATTTTAAAGTGTAAAGGGTTTGAATTTAAATTTGGAG 3782
QY 5536 CTTCAAGACTTTTCTCTATATTTTAAAGTATATTTGTTTATACCAAGACATATCTG 5595
Db GTTTGTTGAGTTATAGTAAATTTAGAGAGTTAAAGTGATTTTCTTATATATAGTATG 3842
QY 5596 AATGTTTATAAATCTTTAAATTTATATATAGTAAATATTTTGTATCAATGCAAT 5655
Db TGTATAGATAGTGTGTTTATTTTGGATTTTATGCGTTTCTTTTGTGTTTATG 3902
QY 5656 ATTTTCT 5696
Db TGTATAGTATTTTGTATTTTATTTTATATATATATATATATATATATATATATATAT 3943

RESULT 6
US-10-204-708-59
; Sequence 59, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE OF INVENTION: By Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10204708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8

NAME/KEY: CDS
LOCATION: (2407) .. (2439)
NAME/KEY: CDS
LOCATION: (2598) .. (3404)
NAME/KEY: CDS
LOCATION: (3580) .. (3720)
NAME/KEY: CDS
LOCATION: (3850) .. (5835)
US-08-213-419B-3

Query Match 0.8%; Score 56; DB 4; Length 6124;
Best Local Similarity 46.8%; Pred. No. 0.0036;
Matches 289; Conservative 0; Mismatches 315; Indels 14; Gaps 3;

Qy 5546 TTTTGGCTTCTATATTTAAGTATATGTTTATAGCAAGAACATATCTGCAATGTTTAT 5605
Db 2640 TTTCTCTGTACATTTTATACATTTTGTGTTAAATATACACCTAAATAAATATGAC 2581
Qy 5606 AAATCTTTAAATTTATATAGTAAATATTTTGTATCACAATGCAATTTTTC 5665
Db 2580 AAATGTTTGTGTAATGCGTAAATAATTTATATTTTATATATATATATTTTAC 2521
Qy 5666 CTCCTTTCCTTCCAACTATACCACTGATTTACCACTTCTAAGAGTACTGACGCGG 5725
Db 2520 ATATTTGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2463
Qy 5726 CCAGATGACCTTCAAGTAGTATATATGATGATATATGATGATATATGATGATAT 5785
Db 2462 TTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 2403
Qy 5786 TACTTCCACTTAACTTCTTAAATTTCTTGGCACTTCCGCAATTTTCAATTTGACACTGG 5845
Db 2402 ATATATGAATATTAAGTTATATATATATTTTAAAGAACTT---TAAATTTTTCGAATATG 2347
Qy 5846 TGTATAGTATAATTTAAATGAATGATATATTTTCTGCAATTTTAAATCTTTTATATGG 5905
Db 2346 TTGATTTTAAATTTCTTTATGTTTAAATATATATATATATATATATATATATAT 2287
Qy 5906 TAGTATTTTATACAGGATATTAACATAGTTAAATCTATGTTTAAATCTTCAATGTTA 5965
Db 2286 TTAATTTTATTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2227
Qy 5966 CAGACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 6025
Db 2226 TAT 2175
Qy 6026 TTTAAGTTAATGTAATAAATTTAATCAATTTTCTGATGCTTTTAAACATTTCTCATGTTAAT 6085
Db 2174 TAAATTTTATTTTATTTTAAATTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 2115
Qy 6086 ATATGTTTGTATCAAAACACTCATATATTTCAAGAAAGAAATATGTTTAAATAGC 6145
Db 2114 TAAATCTAGATATGACCTCTTTTATTTTAAATTTTCAAAAAAATATATATGACACCC 2055
Qy 6146 CCTGTTTAAAGAAATA 6163
Db 2054 TCTCAATAAAAAAATA 2037

RESULT 11
US-08-998-416-1137
Sequence 1137, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
AND USES THEREOF

NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match 0.8%; Score 55.2; DB 3; Length 636;
Best Local Similarity 49.2%; Pred. No. 0.0022;
Matches 174; Conservative 0; Mismatches 178; Indels 2; Gaps 1;

Qy 5327 TGAGGCTTAGACTTATAATCTTTGAATAAACTGATAACTTATTGTATATTTGGAGT 5386
Db 105 TTAATTTTGTATTATTAATCTTATATATATATATATATATATATATATATATAT 164
Qy 5387 GGAGACCTACCTCCATAATAGATAAACTCTTTTGGATATATATCAGAAATTTGCTTT 5446
Db 165 TTATTAATATTTATATAATATAAAATAATATTTAATGAATACTATTATTAGTCTATGT 224
Qy 5447 TTTCTTCTCAAAATTTACATATGATGATATATATATATATATATATATATATATAT 5506
Db 225 TCAAAATTTAATTTAGTTTAAATAATATATATATATATATATATATATATATATAT 284
Qy 5507 TTAATATGATATTAATAAATAATTCGGGCTCTTCAGGACTTTTGTCTTCTATATTTAAGTA 5566
Db 285 TTAATATAGATATATCAATAATTAATAATATATATATATATATATATATATATATAT 342
Qy 5567 TATTGTTTATAGCAAGAACATATTTCTGAATGTTTTTAAATCTTTTAAATATTTTATATG 5626
Db 343 TTTTATATATATAAGATTTAAATTTTAAATATTTGTAATATATATATATATATATATAT 402
Qy 5627 TAGGTAATATTTTGTATCACAATGCAATTTATTTTTTCTCTCTCTCTCTCTCTCT 5680
Db 403 ATCTATTTTATAAATAATTTATGTTGATTTATATATATATATATATATATATATAT 456

RESULT 12
US-08-998-416-288
Sequence 288, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter

```

: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wermland, Jurgen
: APPLICANT: Koechtle, Philipp
: APPLICANT: Rebischung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:

```

	Query Match	0.8%; Score 55.2; DB 3; Length 937;	
	Best Local Similarity	49.2%; Pred. No. 0.0025;	
	Matches 174; Conservative	0; Mismatches 178; Indels	2; Gaps
QY	5327	TGAGGTCTTAGACATTAAATCTTTGGATAAATAAACCTGATTAACCTTATTTGCTATATAAATGGAGT	5386
Ddb	105	TTAATATTATTGAATTAATTAACATTATTATATAAATATTATATAAATTTACTTTAATTCATCA	164
QY	5387	GGAGACCTTACCCTCCATTAATTAGATAAATCTTTTTGGATTATAATCAGAATTTTGCGCTTT	5446
Ddb	165	TTATTAATATTATTATTAATTTAAAAAATAATATTAAATATGNAATCTATTTTAGTCTATAGT	224
QY	5447	TTCTTCTCCAATATTACATATGTATGTATTATATATATCCACATATAFAGTTTTCCCTGA	5506
Ddb	225	TCAAATTTTAAATTAGTTATTAAAATATTATTAGNATATTATTATTCTTTTAAATAAATTA	284
QY	5507	TTAAATGGATATTAANAATAATTCCGGGTGCTTCAGGACCTTTTTGGCTCTATATTATTAGTA	5566
Ddb	295	TTAAATAGATATCAATAAATTAATAATTATTATTATTAAATTTGTTAT--TAAANAATAATATA	342
QY	5567	TATTGTITTTTATAGCAAGAACAATATCTCAATGTTTTTATATAATCTTTAATAAATTTATATG	5626
Ddb	343	TTTTTATTATTATANAGATTTAAATTTTATTAAATATTGTAATATTATTTTTATTATTAAT	402
QY	5627	TAGTAATAATTTTTTGNATCAAGCATTTATTTTTTTTCCCTCTTTTCCTCCAA	5680
Ddb	403	ATCTPATTTTTATAAAATATTATGTTGATTTATATATTATTAAATCTTTTTTATAGAA	456

```

RESULT 13
US-10-204-708-14
; Sequence 14, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: RIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 14
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-14

```

Query Match	0.8%;	Score 53.8;	DB 4;	Length 6113;
Best Local Similarity	43.0%;	Pred. No. 0.012;		
Matches 473;	Conservative 0;	Mismatches 622;	Indels 6;	Gaps 4;
QY	5004	GGAAATGATCTTTTCAAATCATTAGCACAAATATTTAAATATCTAAAAAATTTAAGAGATCCA	5063	
DB	4889	GGAAATGGTTGAGTTAATTTGTTAGGTGATAGTTTAAATTTTTTAAAGAAATGTTTAAAGTGG	4948	
QY	5064	TACTTTCTGTAGCTTTACAAATTAATTTAAAGTACTAAAAAGACAAGGATTTTCTTTTAA	5120	
DB	4949	TGGTATCGTTTATTTTATTTATAGGAGGTATGAGATTTTATAGTTGTTTATATTTTAA	5008	
QY	5121	GAATTTTATAGCAATTTACTGTGTTTATTTAAATGCTAAGCCAAAGATATCTGCACCTTAGGTA	5180	
DB	5009	TTAATATTTGGTATCGGTTAGTTTATTTTAAATTTTATGTTTTTTTAGTAGGTGCTGTCGTA	5068	
QY	5181	TACCTCTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTACAGATGTAACCTTATGAA	5239	
DB	5069	TTTTATTTTTTTTTTTTATGTTTTTTGTTGTTATTAGTTGGTTTTTTTAGATTTTATTTT	5128	
QY	5240	GGAAATATCTGCTTGTGTATATGCGAGTGAAGTACTGGTTCTTCAAAGTCTGCTCAAAAT	5299	
DB	5129	TAATATATAGTAAATTAATTTTAAATTTGGTGATAGTTTATAATTTGAAATATATG	5188	
QY	5300	GTAATTCAGTGGCACAAAAACGATTTTGGAGTCTTAGACTTATAAATCTTTGGAATAAAA	5359	
DB	5189	TGATTTGATGTAATTAATTAATAAATAGGATATAGATAGTTTCGTTATTTTGAATAAAT	5248	
QY	5360	CTGATAACTTATTTGTATAAATTTGGAGTGGAGACCTACCTCCATAATTAGATAAACTCTTT	5419	
DB	5249	TTTTTATGTTATTTTTTTGTAGTTATAAATTTTTTTTTTATTTTAAATTTTAGTAATTTATG	5308	
QY	5420	ATGGAAATATAATCAGAAATTTGCCTTTTTTCTTCTCAAATTAATACATATGTATGATTA	5479	
DB	5309	ATTTGTTTTTGTATATAATTTTTTTTTTATTTGAGATGTTTATATGTATATTTATAT	5368	
QY	5480	TATATCCACATATPAGTTTTTCCCTGATTAATAATGGATATTAATAATATGCGGGTGCTTC	5539	
DB	5369	TATATGTAATTTTTTGGAGTTTGGTTTTTTTTTATTTTAGTATTAATGTTTTTGGAGATTTATA	5428	

5540 AGGACTTTTCTCTATATTAAAGTATATGTTTTATAGCAAGAACATATTTCTGAATG 5599
5429 CGTATATTAGTAGTTGTTTTTTTATTTGTTGTTAGTATTTTATGTTAAATGAGTAT 5488
5600 TTTTATAAATCTTTAATAATTTATATGATGATGATTTTGTATGATGATTTTATTT 5659
5489 TAGTTTCTTTTATTTATTTAAGGGTATTTGAGTTTATGTTATTTTGTAGTTTTCGTA 5548
5660 TTTTCTCTCTCTCTCCAACTATACCACTGTAATTTACCACTTCTTAAGAGTGAATGAC 5719
5549 TTATGAATATAGTTGTAAGAAATATTTATATATGCTTCTGTTGAATATAGTTTAT 5608
5720 GACGGCCAGATGACCTTGAAGTAGTATTTATGATGATGATGATGATGATGATGATG 5779
5609 TTTTATGAGTATATATTTAAGAGTGTGAGTGTGTTGTTGTTATAGTATGTTTATTT 5668
5780 TTTTCTTCT 5839
5669 GTAAACCGGAATAATATGATTTATTTAGTTTATTAAGATTTTATTAAGATTTTAT 5728
5840 CACTGGTGTATAGTATATAAATTTAAATGAATTAATTAATTAATTAATTAATTAATTT 5899
5729 ATAAATTTGATTAACGTATTTTAAAGAGATGAGATTTAAATGTTTAAATTTAGTT 5788
5800 ATATGAGTATTTTATTAACAGGATTTTAAATTAATTAATTAATTAATTAATTTGAA 5958
5789 AATTTATAGATTTTATTTGTTTATTTTAAATTTTATTAATTTTATTAATTTTAT 5848
5959 ATGTTACAGAGCTTCT 6018
5849 AGTATAGATTTATGATGATTTTGTAAATTTATTAATTAATTAATTAATTTATTT 5908
6019 CCGTCAATTTAAGTATGTAATAAATTTAATTAATTTATTAATTTTATTAATTTTAT 6078
5909 TATTATAATGA-TGTTTAAACGTAATTTTAAATTTTATTAATTTTATTAATTTGAA 5967
6079 GTGTAATATATGTTTCTGAT 6099
5968 GTATAATAGATTTTATTAATG 5988

RESULT 14
US-10-204-708-60
; Sequence 60, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 60
; LENGTH: 6156
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-60

Query Match 0.8%; Score 53.6; DB 4; Length 6156;
Best Local Similarity 42.7%; Pred. No. 0.013;
Matches 466; Conservative 0; Mismatches 609; Indels 17; Gaps 3;
QY 5035 TTTTAAATCTAAAAATTTAAGAGATCCATCTTCTCTAGCTTTTACAAATTAATTTAAGT 5094
Db 3335 TTGTTTATAGTATATTTTATGTTAGTCTGATTTTAGGAAGTTTATTTATGTTATGT 3394
QY 5095 ACTAAAAAGACAAGGATTTCTTTTAAAGAAATTTATAGCATTTTACGTGTTATTTAAATGC 5154
Db 3395 TATATATAGATTTATATAATTTTACGGGAATTTTAAATTTTATAGAGTTAATAGTTTA 3454
QY 5155 TAAGCCAAAGTATCTGCATTTAGGTATACCTCTTATGCCAATATGATTTTAAATGAAG 5214
Db 3455 GTTGTTTTAAATTTTATGATTTTAAAGATTTTATGTTATGATATATAATTTTGAAGAAAG 3514
QY 5215 CTCCTTTTCAGATGTAACCTTTATGAAGAAATATCTGCTTTTGTATATATGCGAGTTAGAT 5274
Db 3515 ATATTTTAAATTTTAAATTTAAGATTTTATGTTATGATATATAATTTTGAAGAAAGAT 3574
QY 5275 ACTGGTTTCTAAAGTCTGTCAAAATTTGATTTTTCAGTGGCACAACAAACAGTTTGGTCT 5334
Db 3575 ATATATTTTAAATTTAAGAGTGTGATTTTCTTCTTCTCTCTCTCTCTCTCTCTCTCT 3625
QY 5335 TAGACTTATAATTTCTTTGAATTAATACTGATACTTATTTGTATATAATTTGGAGTGGAGCT 5394
Db 3626 TTTTATTAAGAGTGTGTTAAAGTTTATTTGAGGTAGTTATTTATAGAAATAGAAATTTGAT 3685
QY 5395 ACCTCATTAATTTAGATAAACTCTTTTGGATTTATTAATCAGAATTTTGCCTTTTCTCTCT 5454
Db 3686 TTCGGTATATAAATATTTATTTTAAAGATTAAGTAGACGTATATTTGTTGTTGAAT 3745
QY 5455 CAAATTTATACATATGATGATATATATATCCACATAT-ATAGTTTCTCTCTCTCTCTCTCT 5513
Db 3746 AATGGAATATTTTAAAGATTTAAATTTTATTTATTTAAATTTTAAATTTAAATTAAT 3805
QY 5514 GATATTTAAATTAATTCGGGTCTTCAGGACTTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 5573
Db 3806 GTTATAATTTTAAAGATTTATAGTATTTTAAATTTTAAATTTTAAATTTTAAATTTT 3865
QY 5574 TTTTACGCAAGACATATTTCTGAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 5633
Db 3866 ATATTTTATAGTTTATATAAAGTTATGAGGAGATTTTAAATTTTAAATTTTAAATTTT 3925
QY 5634 TATTTTGTATCACAATGCAATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5693
Db 3926 GTTATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 3985
QY 5694 ATTTACACTTCTTAAGAGTGAATGACGACGGCCAGATGACCTTGAAGTATGATGATG 5753
Db 3986 TTTTAGCAATTTATTTTATGTTATTTAGTGGGAGTTTATTTAAATTTTAAATTTTAAAT 4045
QY 5754 TAGCAATAAATGAAGCTGAACAGGTTTTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5806
Db 4046 GAAATTTTGGTGGTGTAGTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 4105
QY 5807 AAAATTTCTTGGCAACTTCCGATATTTTATGATGACCTGTTGATATAGTATATAATTTAAAT 5866
Db 4106 AAATATATTATGATTTAAGTGTATATTTTATGTTTATTTTAAATTTTAAATTTTAAAT 4165
QY 5867 GAATAATTTACTTTTGCATTTTAAATTTCTTTATGTTAGTATTTTAAATTTTAAATTTTAA 5926
Db 4166 GGGTGGTGTATTAATTTAATTTGATATAAATAATTTTGTATATAATTTTAAATTTTAAAT 4225
QY 5927 TATTAACATAAAGTTAAATCTCTATGATTTTGAATTTGTAAGATTTGTAAGAGCTTTCTCTCT 5986
Db 4226 TTTTGTAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4285
QY 5987 AAACAGCAAAAAAGTGGGGGCAATTTGTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6046
Db 4286 AGGTGTTTAAATTTTGGAGGGGGTGGGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4345

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2004, 02:10:43 ; Search time 1709.6 Seconds
(without alignments)
17604.315 Million cell updates/sec

Title: US-09-744-167-3

Perfect score: 5632

Sequence: 1 attccggccgggtagctc.....ttggaggcacattttgaagt 6632

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubnpa/US05_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubnpa/US05_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq*
- 8: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq*
- 9: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubnpa/US09_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubnpa/US09_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubnpa/US10C_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq*
- 18: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5290	79.8	5402	15	US-10-072-977-11
2	1514	22.8	1572	9	US-09-925-299-51
3	1514	22.8	1572	10	US-09-925-299-51
4	850	12.8	969	16	US-10-264-237-394
5	472.4	7.1	475	10	US-09-918-995-3724
6	444	6.7	455	10	US-09-930-213-158
7	389.4	5.9	391	10	US-09-918-995-7316
8	300.4	4.5	420	13	US-10-085-783A-29775
9	300.4	4.5	420	16	US-10-242-535A-29775
10	286	4.3	317	15	US-10-106-698-2133
11	199.6	3.0	393	9	US-09-960-352-13367
12	181.6	2.7	289	14	US-10-040-739-971
13	172	2.6	556	9	US-09-864-761-13497
14	168	2.5	168	9	US-09-864-761-30043
					Sequence 11, Appl
					Sequence 51, Appl
					Sequence 51, Appl
					Sequence 394, Appl
					Sequence 3724, Ap
					Sequence 158, Ap
					Sequence 7316, Ap
					Sequence 29775, A
					Sequence 29775, A
					Sequence 2133, Ap
					Sequence 13367, A
					Sequence 971, Ap
					Sequence 13497, A
					Sequence 30043, A

C 15	124	1.9	387	9	US-09-880-107-5	Sequence 5, Appl
C 16	93.6	1.4	255	10	US-09-930-213-577	Sequence 577, Appl
C 17	88.8	1.3	3673778	15	US-10-312-841-1	Sequence 1, Appl
18	85.4	1.3	474	13	US-10-027-632-195942	Sequence 195942,
19	85.4	1.3	474	16	US-10-027-632-195942	Sequence 1721, Ap
20	82	1.2	11805	15	US-10-311-455-1721	GENERAL INFORMATI
C 21	77.2	1.2	158001	17	US-10-211-179-11	Sequence 986, App
22	75.6	1.1	7597	15	US-10-311-455-986	Sequence 185, App
23	75.2	1.1	7597	15	US-10-239-676-185	Sequence 185, App
24	75.2	1.1	7597	15	US-10-311-455-1995	Sequence 1995, Ap
25	75	1.1	11812	13	US-10-231-714A-468	Sequence 468, App
26	75	1.1	11812	15	US-10-239-676-210	Sequence 210, App
27	75	1.1	11812	15	US-10-311-455-2092	Sequence 2092, Ap
28	75	1.1	11812	15	US-10-240-453-306	Sequence 306, App
29	74	1.1	61020	13	US-10-221-714A-514	Sequence 514, App
30	73.2	1.1	5689	13	US-10-221-714A-148	Sequence 148, App
31	73.2	1.1	5689	15	US-10-239-676-90	Sequence 90, Appl
32	73.2	1.1	5689	15	US-10-240-453-100	Sequence 100, App
33	73	1.1	5920	15	US-10-311-455-490	Sequence 490, App
34	72.6	1.1	5527	15	US-10-311-455-290	Sequence 290, App
35	72.6	1.1	5527	15	US-10-240-452-38	Sequence 38, Appl
36	72.4	1.1	15548	15	US-10-311-455-2128	Sequence 2128, Ap
37	72.2	1.1	15399	15	US-10-311-455-1488	Sequence 1488, Ap
38	71.2	1.1	6681	15	US-10-311-455-128	Sequence 128, App
39	71.2	1.1	6681	15	US-10-240-452-4	Sequence 4, Appl
40	71.2	1.1	18218	15	US-10-311-455-1922	Sequence 1922, Ap
41	71	1.1	6112	15	US-10-311-455-462	Sequence 462, App
42	71	1.1	7498	15	US-10-311-455-230	Sequence 230, App
43	70.8	1.1	6314	15	US-10-240-452-14	Sequence 14, Appl
44	70.6	1.1	17183	15	US-10-311-455-459	Sequence 459, App
45	70.2	1.1	15732	15	US-10-239-676-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1

US-10-072-977-11

; Sequence 11, Application US/10072977

; Publication No. US20030152933A1

; GENERAL INFORMATION:

; APPLICANT: Barash et al.

; TITLE OF INVENTION: Human Secreted Proteins

; FILE REFERENCE: PT057P1

; CURRENT APPLICATION NUMBER: US/10/072,977

; CURRENT FILING DATE: 2002-02-12

; PRIOR APPLICATION NUMBER: PCT/US01/25288

; PRIOR FILING DATE: 2001-08-13

; PRIOR APPLICATION NUMBER: US 60/225,215

; PRIOR FILING DATE: 2000-08-14

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 11

; LENGTH: 5402

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-072-977-11

Query Match 79.8%; Score 5290; DB 15; Length 5402;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 5317; Conservative 0; Mismatches 20; Indels 5; Gaps 1;

QY	185	ATATCTCTCCAGGATGTTCTCAAGGCATACAGAATTAATTCGTAATTAAGTCTGCAGG	244
DB	48	AAAGCTCCGAGGGCTGTAGGAGGAGGATACAGAATTAATTCGTAATTAAGTCTGCAGG	107
QY	245	TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCGACAGTTCGACAACTCCTTGAGTATTT	304
DB	108	TAGGATGACAGTATTTTAAAGCAGCTGTGAGTCGACAGTTCGACAACTCCTTGAGTATTT	167
QY	305	TGAACAGAACCCAGATGAACAGAGATTATCTCGAGATGTACAAATGCATATGATTTCTAA	364
DB	168	TGAACAGAACCCAGATGAACAGAGATTATCTCGAGATGTACAAATGCATATGATTTCTAA	227

QY 365 CCACTGCTCAGTTTCTTCCAGAGTGGCTTCCCTCAGACGGAATCTCATTGCTCCCAAGA 424
Db 228 CCACTGCTCAGTTTCTTCCAGAGTGGCTTCCCTCAGACGGAATCTCATTGCTCCCAAGA 287
QY 425 CCAAGAGTGGTTAATAGTTGTGCTCATCAGAAACAGCTATGGAACAATAGAGATTC 484
Db 288 CCAAGAGTGGTTAATAGTTGTGCTCATCAGAAACAGCTATGGAACAATAGAGATTC 347
QY 485 CTTGAATGAAAAACAACCTCAAGGGACTTACTTCTATACAAAATGAAAAAATGTAACAGG 544
Db 348 CTTGAATGAAAAACAACCTCAAGGGACTTACTTCTATACAAAATGAAAAAATGTAACAGG 407
QY 545 ACTGATCTCTTCTTCTGCTGATGGTGGTACTTCCAGATGAATCCAGCGCTTATATAT 604
Db 408 ACTGATCTCTTCTTCTGCTGATGGTGGTACTTCCAGATGAATCCAGCGCTTATATAT 467
QY 605 GGGACGATCTAGTAAACCTATCTGTGATCTGATAGTACATGGTAACTTAGTTCATGC 664
Db 468 GGGACGATCTAGTAAACCTATCTGTGATCTGATAGTACATGGTAACTTAGTTCATGC 527
QY 665 AACCAATAGTGAAGAGATATTAAAAAATTTAAGCCAGATGATTTTAAAGTCTAATGCAGA 724
Db 528 AACCAATAGTGAAGAGATATTAAAAAATTTAAGCCAGATGATTTTAAAGTCTAATGCAGA 587
QY 725 TTCCCTGATGATGGATTTATCTTCACTGTCAGATCTCCCTGCTGTTCTTCAACAGA 784
Db 588 TTCCCTGATGATGGATTTATCTTCACTGTCAGATCTCCCTGCTGTTCTTCAACAGA 647
QY 785 CCATGATAGTACTGTCAGAGAACACAGAAATGATATCAGTCTGAAATACAAAATAG 844
Db 648 CCATGATAGTACTGTCAGAGAACACAGAAATGATATCAGTCTGAAATACAAAATAG 707
QY 845 AGAAATCGAGGAATCAAGAAATGGGTATAAAGTAGATACACACTTCAGATTCCTTA 904
Db 708 AGAAATCGAGGAATCAAGAAATGGGTATAAAGTAGATACACACTTCAGATTCCTTA 767
QY 905 TAAATACAGTGAACAGAAAAATTTAAAAAGATAAAAGATCTTTAATCAGTTAGAAATCAAT 964
Db 768 TAAATACAGTGAACAGAAAAATTTAAAAAGATAAAAGATCTTTAATCAGTTAGAAATCAAT 827
QY 965 TGTGATTTAAACATGTCATCTGCTTTCAGTCTGACGAAAGTTCCAAAATGTTTCATGCCAA 1024
Db 828 TGTGATTTAAACATGTCATCTGCTTTCAGTCTGACGAAAGTTCCAAAATGTTTCATGCCAA 887
QY 1025 AGACAAGCTACAAACAGAGCCAGCCATGGGATTAATAAGATCTTGGCTTAGTAAA 1084
Db 888 AGACAAGCTACAAACAGAGCCAGCCATGGGATTAATAAGATCTTGGCTTAGTAAA 947
QY 1085 AGAGAAGTAGATGTGGCAGTCATAACTCGCGCAGAAATGTTTAAAGAGAGGGCAGAGAC 1144
Db 948 AGAGAAGTAGATGTGGCAGTCATAACTCGCGCAGAAATGTTTAAAGAGAGGGCAGAGAC 1007
QY 1145 AAGTCTTTTACCTGACGCTTCGAAAAATGAAGATTTATGCTTAAATGATTCAAATTC 1204
Db 1008 AAGTCTTTTACCTGACGCTTCGAAAAATGAAGATTTATGCTTAAATGATTCAAATTC 1067
QY 1205 AAGAGATGAAATTTCAAATTTACCTGACTTTTCCCTTTCAGGAAGATAAGACTGTTATAAA 1264
Db 1068 AAGAGATGAAATTTCAAATTTACCTGACTTTTCCCTTTCAGGAAGATAAGACTGTTATAAA 1127
QY 1265 ACAATCTGCAACAAGACTCAAAAATTTAGACCTTAAGATAATGATGTAATCCAGA 1324
Db 1128 ACAATCTGCAACAAGACTCAAAAATTTAGACCTTAAGATAATGATGTAATCCAGA 1187
QY 1325 TTCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCTCTCATTTGCTGCTTCC 1384
Db 1188 TTCTCTTTCAGCTTTTACATGTTTCCAGTAAAGATGTGCGCTCTCTCATTTGCTGCTTCC 1247
QY 1385 TGGCTGTGGGTCTATGTGTGATCATTTAATGAAAGTAAAGCAAGGGGTGATTTTATACC 1444
Db 1248 TGGCTGTGGGTCTATGTGTGATCATTTAATGAAAGTAAAGCAAGGGGTGATTTTATACC 1307

QY 1445 TCAGCTGAACATAAAGATAATATACAAGATGCAGTGACTATACATGAAGAAATACAGAA 1504
Db 1308 TCAGCTGAACATAAAGATAATATACAAGATGCAGTGACTATACATGAAGAAATACAGAA 1367
QY 1505 CAGTGTGTTCTAGTGGGGAACCATTCAAAGAGATGATCTTTTGAACAGGAAAAATG 1564
Db 1368 CAGTGTGTTCTAGTGGGGAACCATTCAAAGAGATGATCTTTTGAACAGGAAAAATG 1427
QY 1565 TAAAGCATACTCTCTCAGTCATTAAATGAAGGGATGGAAGACAGAAAGATAGTCTCTCA 1624
Db 1428 TAAAGCATACTCTCTCAGTCATTAAATGAAGGGATGGAAGACAGAAAGATAGTCTCTCA 1487
QY 1625 CCAGACAGTAATCAGAGCTGAGTCTTTTGGATGGTGTGACACAGTCTTACAGTTGTAGA 1684
Db 1488 CCAGACAGTAATCAGAGCTGAGTCTTTTGGATGGTGTGACACAGTCTTACAGTTGTAGA 1547
QY 1685 ATCTCAAGAGGGCTTCTTGGGCACTCATGTCCAGAGTCTTCTGATTTGTGTGAGGTTT 1744
Db 1548 ATCTCAAGAGGGCTTCTTGGGCACTCATGTCCAGAGTCTTCTGATTTGTGTGAGGTTT 1607
QY 1745 TATTAATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTTACTTTAATTTGA 1804
Db 1608 TATTAATACTTTTTCAAGCAATGATATGGATGGGCAAGACTTAGATTTACTTTAATTTGA 1667
QY 1805 TGAAGCGCAAAAAGTGGGCCACTTAATAGTATGATGCTGAACCTTGATGCCCTTCTCAGAGA 1864
Db 1668 TGAAGCGCAAAAAGTGGGCCACTTAATAGTATGATGCTGAACCTTGATGCCCTTCTCAGAGA 1727
QY 1865 ACAGTATCTTTCAGACCCTAAACATAAAGTCTTTTGAAGAAAAATGTAATGACTCTAAATC 1924
Db 1728 ACAGTATCTTTCAGACCCTAAACATAAAGTCTTTTGAAGAAAAATGTAATGACTCTAAATC 1787
QY 1925 GCAATGAATCAGATAGATATGAAGGGCTTAGATGATGGAACACATCAATATATATATTT 1984
Db 1788 GCAATGAATCAGATAGATATGAAGGGCTTAGATGATGGAACACATCAATATATATATTT 1847
QY 1985 CAATGCAAGCAGAGAGTATTTGGGGAAGTCAATGATTAATAATAATTTGTGAAACAGT 2044
Db 1848 CAATGCAAGCAGAGAGTATTTGGGGAAGTCAATGATTAATAATAATTTGTGAAACAGT 1907
QY 2045 TGATAAACAAAAATACAAATAGAAAAATGGCTTTCTTTAGGAGAAAAAGACACTATTTCCAGT 2104
Db 1908 TGATAAACAAAAATACAAATAGAAAAATGGCTTTCTTTAGGAGAAAAAGACACTATTTCCAGT 1967
QY 2105 TCACAAGGGTTTACCTACAGTACAGTCTGAGATTAACAATCAATTTACGCTCTCTGATAT 2164
Db 1968 TCACAAGGGTTTACCTACAGTACAGTCTGAGATTAACAATCAATTTACGCTCTCTGATAT 2027
QY 2165 TTAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2224
Db 2028 TTAACAGTCAATCTGTTGGAGGGCCAGACCTTAAGCAATTTGTTTACCTTCCATCAAGAAC 2087
QY 2225 AAGGAGTTCAAGAGGACCTGAAATAGCCAGATGTTCCAGATACAAATAGAAAGTGAACCCAG 2284
Db 2088 AAGGAGTTCAAGAGGACCTGAAATAGCCAGATGTTCCAGATACAAATAGAAAGTGAACCCAG 2147
QY 2285 CACAGCAGATACCGTGTGTTTCCAAATCACTTGTGCTATAGATTTCTACAGCTGATCCACAGGT 2344
Db 2148 CACAGCAGATACCGTGTGTTTCCAAATCACTTGTGCTATAGATTTCTACAGCTGATCCACAGGT 2207
QY 2345 TAGCTTCAACTCTAATATCATTTGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTTCGT 2404
Db 2208 TAGCTTCAACTCTAATATCATTTGATATAGAAAGTAATTTCTGAAGTGGATCTAGTTTCGT 2267
QY 2405 AACTGCAATGAAGATTTCTGACCTGAAAAACACTTGCAAGAAAGCTTGGTTTGGGCCA 2464
Db 2268 AACTGCAATGAAGATTTCTGACCTGAAAAACACTTGCAAGAAAGCTTGGTTTGGGCCA 2327
QY 2465 GAAACAGCCTACTTGGGTTCTCTGATTCAGAAAGTCCAAACTGTATGAACTGCCAAGTCAA 2524
Db 2328 GAAACAGCCTACTTGGGTTCTCTGATTCAGAAAGTCCAAACTGTATGAACTGCCAAGTCAA 2387
QY 2525 ATTTACTTTTTTACCAACCGGCGACACCACTTGCCGAGCATGTGGGAAAGTATTTTCTGGTGT 2584


```
QY 5705 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATAAAT 5764
DB 840 CTAAGAGTGACTGACGACGGCCAGATGACCCCTGAAGTAGTCATTATGTAGCAATAAAT 899
QY 5765 GAAGCCTGAACACAGGTTTTTACTTTCACCTTTAAATCCTTAGAAATTTCTTGGCAACTTC 5824
DB 900 GAAGCCTGAACACAGGTTTTTACTTTCACCTTTAAATCCTTAGAAATTTCTTGGCAACTTC 959
QY 5825 GCATATTTCTTATGACACGTGTGTATAGTATAATTTAAATGAACATAATTTACTTTTGGCA 5884
DB 960 GCATATTTCTTATGACACGGGTGTATAGTATAATTTAAATGAACATAATTTACTTTTGGCA 1019
QY 5885 TATTTTAAATTTCTTATATGTTAGTTATTTTATACAGGATATTAACATAAGCTTAAAT 5944
DB 1020 TATTTTAAATTTCTTATATGTTAGTTATTTTATACAGGATATTAACATAAGCTTAAAT 1079
QY 5945 CCTATGATTTTGAATTTCTTACAGAGCTTTCTCTTTACTTCAACACGCAAAAAAGTGGG 6004
DB 1080 CCTATGATTTTGAATTTCTTACAGAGCTTTCTCTTTACTTCAACACGCAAAAAAGTGGG 1139
QY 6005 GGGCATATTTAGTCTCTGCTCAATTTAAGTTATGTAAATAATTTAATCATTTATTTGATGCT 6064
DB 1140 GGGCATATTTAGTCTCTGCTCAATTTAAGTTATGTAAATAATTTAATCATTTATTTGATGCT 1199
QY 6065 TTAAACATTTCTCATGTGTAATATATGTTTTTGTATCAAAAAACACTCATATATTTCAAGAA 6124
DB 1200 TTAAACATTTCTCATGTGTAATATATGTTTTTGTATCAAAAAACACTCATATATTTCAAGAA 1259
QY 6125 AAAGAAATTTGTTAAATAGCCCTGTTTTTAAGAAATAATTTATGAAGCATCTCAACTTG 6184
DB 1260 AAAGAAATTTGTTAAATAGCCCTGTTTTTAAGAAATAATTTATGAAGCATCTCAACTTG 1319
QY 6185 AAGATCAAGTCAAGTTTAACTCAGGATCTGAGGTCTCAAGCTAGGAGACCTGAGAAAT 6244
DB 1320 AAGATCAAGTCAAGTTTAACTCAGGATCTGAGGTCTCAAGCTAGGAGACCTGAGAAAT 1379
QY 6245 TTTAATCAGTTTGGCATATATAGTTTGGACTGAAATCAATCTGTAGTACTTAGGCCAAGAC 6304
DB 1380 TTTAATCAGTTTGGCATATATAGTTTGGACTGAAATCAATCTGTAGTACTTAGGCCAAGAC 1439
QY 6305 AATTTGGAGGAGAAATATCAGCCCTCTGGAAGTACTTCTGGAACAAATGTAAAGTCTC 6364
DB 1440 AATTTGGAGGAGAAATATCAGCCCTCTGGAAGTACTTCTGGAAGTACTTCTGGAAGTCTC 1499
QY 6365 GCAGATATTTCAATAAAATGGCAACCTGTTTAA 6397
DB 1500 GCAGATATTTCAATAAAATGGCAACCTGTTTAA 1532

RESULT 3
US-09-925-299-51
; Sequence 51, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1555)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-51
```

```
Query Match 22.8%; Score 1514; DB 10; Length 1572;
Best Local Similarity 99.3%; Pred. No. 6.2e-304;
Matches 1522; Conservative 7; Mismatches 3; Indels 1; Gaps 1;
```

```
QY 4865 TTAGTGAAGAATGTGCCATATATACATATTGCAACCTAATTTGTAAAACTAACTCCAGC 4924
DB 1 TNAGTGAAGAATGTGCCATATATACATATTGCAACCTAATTTGTAAAACTAACTCCAGC 60
QY 4925 ACTAAAGCTCAAAATGCCACAAAACACATAAAGTATAATATGCTGATTTTGAACACAT 4984
DB 61 ACTAAAGCTCAAAATGCCACAAAACACATAAAGTATAATATGCTGATTTTGAACACAT 120
QY 4985 AAGCTTTGCTCTTTAGGAGGAATGATCTTTCAATCATATTAGCAACAATTTAAATATC 5044
DB 121 AAGCTTTGCTCTTTAGGAGGAATGATCTTTCAATCATATTAGCAACAATTTAAATATC 180
QY 5045 TAAAAATTTAAGAGATCCATACCTTTCTGTAGCTTTACAAATTAATTTAAGTACTTAAAAA 5104
DB 181 TAAAAATTTAAGAGATCCATACCTTTCTGTAGCTTTACAAATTAATTTAAGTACTTAAAAA 240
QY 5105 CAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCCAAAG 5164
DB 241 CAAGGATTTCTTTTAAAGAAATTTATAGCAATTTACTGTGTTATTTAAATGCTAAGCCAAAG 300
QY 5165 TATCTGCACCTTAGGTATACCTCTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAG 5224
DB 301 TATCTGCACCTTAGGTATACCTCTTTATGCCAATAATGATTTTAAATGAAGGCTCTTTTCAG 360
QY 5225 ATGTAACTTTATGAAGAAATATCTGCTTTGTGTATATGCGAGTTAGAATACCTGTTTCT 5284
DB 361 ATGTAACTTTATGAAGAAATATCTGCTTTGTGTATATGCGAGTTAGAATACCTGTTTCT 420
QY 5285 AAAGTCTGCAAAATGTATTTTCAAGTGCACAAAAACAGTTTTCAGTCTTAGACTTATA 5344
DB 421 AAAGTCTGCAAAATGTATTTTCAAGTGCACAAAAACAGTTTTCAGTCTTAGACTTATA 480
QY 5345 ATCTTTGAATAAAATCTGATAAATTTTGTATAAATTTGAGTGGAGACCTACCTCCATAA 5404
DB 481 ATCTTTGAATAAAATCTGATAAATTTTGTATAAATTTGAGTGGAGACCTACCTCCATAA 540
QY 5405 TTAGATAAACTCTTTTGGATATAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 5464
DB 541 TTAGATAAACTCTTTTGGATATAATCAGAAATTTTGCCTTTTCTCTCAAAATTTATA 600
QY 5465 CATATGTATGTATATATATATCCATATATAGTTTTCCTGATTAAATGGATATTAAT 5524
DB 601 CATATGTATGTATATATATATCCATATATAGTTTTCCTGATTAAATGGATATTAAT 660
QY 5525 AATTGCGGGTCTTCAGGACTTTTGTCTATATTTTAAAGTATAATGTTTTTATAGCAAG 5584
DB 661 AATTGCGGGTCTTCAGGACTTTTGTCTATATTTTAAAGTATAATGTTTTTATAGCAAG 720
QY 5585 AACATATCTGAATGTTTTTATAAATCTTTAATTTATATGTTAGTAAATATTTTCTAT 5644
DB 721 AACATATCTGAATGTTTTTATAAATCTTTAATTTATATGTTAGTAAATATTTTCTAT 780
QY 5645 CACAATGCAATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5704
DB 781 CACAATGCAATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 839
QY 5705 CTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTATGTAGCAATAAAT 5764
DB 840 CTAAGAGTACTGACGACGGCCAGATGACCCCTTGAAGTAGTCATTATGTAGCAATAAAT 899
QY 5765 GAAGCCTGAACACAGGTTTTTACTTTCACCTTTAATCCTTAGAAATTTCTTGGCAACTTC 5824
DB 900 GAAGCCTGAACACAGGTTTTTACTTTCACCTTTAATCCTTAGAAATTTCTTGGCAACTTC 959
```

Qy	5825	GCATATTTTTCATTGACACTGGTGATTAAGTATAAAATTTAAATGAACATAATCTTTTGCA	5884
Db	960	GCATATTTTTCATTGACACCGGTGATTAAGTATAAAATTTAAATGAACATAATCTTTTGCA	1019
Qy	5885	TATTTTAAATTCCTTATATGGTAGTATTTTATTAACAGGATATTAAACATAAGTTAAAT	5944
Db	1020	TATTTTAAATTCCTTATATGGTAGTATTTTATTAACAGGATATTAAACATAAGTTAAAT	1079
Qy	5945	CCATGATTTTGAATTTGTACAGAGCTTCCCTCTTTTCAACACAGCAAAAAGTGGG	6004
Db	1080	CCATGATTTTGAATTTGTACAGAGCTTCCCTCTTTTCAACACAGCAAAAAGTGGG	1139
Qy	6005	GGGCATATTCGAGTCGTCATTTAGTTATGTAAAAAATTTAATCATTTATTTGATGCT	6064
Db	1140	GGGCATATTCGAGTCGTCATTTAGTTATGTAAAAAATTTAATCATTTATTTGATGCT	1199
Qy	6065	TTAAACATTCCTCATGTGTAATATATGTTTTGTTGATCAAAAACACTCATATATTTCAAGAA	6124
Db	1200	TTAAACATTCCTCATGTGTAATATATGTTTTGTTGATCAAAAACACTCATATATTTCAAGAA	1259
Qy	6125	AAAGAAATTTATGTTAAATAGCCCTGTTTTAAGAAAAATTTTATGAAGCATCTCAACTTG	6184
Db	1260	AAAGAAATTTATGTTAAATAGCCCTGTTTTAAGAAAAATTTTATGAAGCATCTCAACTTG	1319
Qy	6185	AAGATCAAGTCAAAGTTATAACTCAGGATCTGAGTGCTCAAGCTAGGAGAGACTGAGAAT	6244
Db	1320	AAGATCAAGTCAAAGTTATAACTCAGGATCTGAGTGCTCAAGCTAGGAGAGACTGAGAAT	1379
Qy	6245	TTTAAATCAGHTTGGGCATATAGTTTGGACTGAATTCACATCTGTPAGTACTTTAGCCAAAGAC	6304
Db	1380	TTTAAATCAGHTTGGGCATATATTTGGACTTGAATTCACATCTGTPAGTACTTTAGCCAAAGAC	1439
Qy	6305	AATTTGGAGAGAAATATCAGCCCTTCTGGAAAGTAGTACTTTCCTGGAACAATGTAAAGTGTC	6364
Db	1440	AATTTGGAGAGAAATATCAGCCCTTCTGGAAAGTAGTACTTTCCTGGAACAATGTAAAGTGTC	1499
Qy	6365	GCAGATATTTCAATAAAAATGGCAACCTGTTATAA	6397
Db	1500	GCARATATTTCAATAAAAATGGCAACCTGTTAAAA	1532

RESULT 4

```

US-10-264-237-394
; Sequence 394, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 394
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (845)..(845)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871)..(871)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (897)..(897)

```


Db 897 NTNTGAAATGGAAAAACCGCTTAAATTTCCCGGAAAA 938

RESULT 5

US-09-918-995-3724
; Sequence 3724, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3724

Query Match 7.1%; Score 472.4; DB 10; Length 475;

Best Local Similarity 99.8%; Pred. No. 8e-88;

Matches 473; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3496 ACTTTTCAACCAATGGATGCTTGGGACGAGCAATATTATTCTATTGTTAT 3555

Db 1 ACTTTTCAACCAATGGATGCTAGGACGAGCAATATTATTCTATTGTTAT 60

Qy 3556 GTTTCGCAATGAAGATCTATTCTTAAGGACATCTTCAGACTATTATACCAATATA 3615

Db 61 GTTTCGCAATGAAGATCTATTCTTAAGGACATCTTCAGACTATTATACCAATATA 120

Qy 3616 AGGATGCTCTAAAGGAAATACATAGAAACTTCGACAAATATTACCTTTACTGAGATT 3675

Db 121 AGGATGCTCTAAAGGAAATACATAGAAACTTCGACAAATATTACCTTTACTGAGATT 180

Qy 3676 TTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACCTACTTTTCAGAACTTG 3735

Db 181 TTCTCAGTAGCAAGGATCAGGAGGATTCCTGTTTATTACCTACTTTTCAGAACTTG 240

Qy 3736 ATGATCTCTCATACAGTAATCTTTCTTTGTTGGAATCTTATCCAGAGCTTGAGA 3795

Db 241 ATGATCTCTCATACAGTAATCTTTCTTTGTTGGAATCTTATCCAGAGCTTGAGA 300

Qy 3796 TTCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGGGTGCAGAAATATAAG 3855

Db 301 TTCCCTGGGCAAGGTTTTCCTATGCGTTTAAATGTTGAGATTGGGTGCAGAAATATAAG 360

Qy 3856 CATATCTCTCTCTCTAACAAGCATCAGAGCGCGAAACCTCTTTTTCGAGAAATAGGAC 3915

Db 361 CATATCTCTCTCTCTAACAAGCATCAGAGCGCGAAACCTCTTTTTCGAGAAATAGGAC 420

Qy 3916 ACATATTATGAACTTACTTGTGTTGACCTTCGAAATTTACAGTATACCTTTGCATA 3969

Db 421 ACATATTATGAACTTACTTGTGTTGACCTTCGAAATTTACAGTATACCTTTGCATA 474

RESULT 6

US-09-930-213-158/c
; Sequence 158, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHE-NITSE, OLEG
; APPLICANT: GRIFS, MARTIN
; APPLICANT: HELLMER, MARTIN

; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SERS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBRE-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 885
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 158
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-930-213-158

Query Match 6.7%; Score 444; DB 10; Length 455;

Best Local Similarity 99.8%; Pred. No. 6.2e-82;

Matches 455; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 5337 GACTTATAATCTTTGAATAAACTGATAACTTATTGTATATAATGGAGTGGAGACTAC 5396

Db 455 GACTTATAATCTTTGAATAAACTGATAACTTATTGTATATAATGGAGTGGAGACTAC 396

Qy 5397 CTCATTAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTGCTTTTCTTCTCA 5456

Db 395 CTCATTAATTAGATAAACTCTTTTGGATTATAATCAGAAATTTTGCTTTTCTTCTCA 336

Qy 5457 AATTATTACATATATGATGATTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 5516

Db 335 AATTATTACATATATGATGATTATATATATCCACATATATAGTTTCCCTGATTAATGGAT 276

Qy 5517 ATTAAATAATTCGGGTGCTTCAGGACCTTTTGGCTTCTATATTAAGTATATTTT 5576

Db 275 ATTAAATAATTCGGGTGCTTCAGGACCTTTTGGCTTCTATATTAAGTATATTTT 216

Qy 5577 ATAGCAAGACATATCTGAATGTTTATAAACTTTATAAATTTATATAGTAAATAT 5636

Db 215 ATAGCAAGACATATCTGAATGTTTATAAATTTATAAATTTATATAGTAAATAT 156

Qy 5637 TTTTGTATCAATGCAATATTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGATT 5696

Db 155 TTTTGTATCAATGCAATTA-TTTTTCCTCTCTCTTCCCTTCCAACTATACCACTGATT 97

Qy 5697 TACCACCTCTAAGAGTGAATGACGCGGCGGACGATGACCTTGAAGTAGTATATCTAG 5756

Db 96 TACCACCTCTAAGAGTGAATGACGCGGCGGACGATGACCTTGAAGTAGTATATCTAG 37

Qy 5757 CAATAAATGAAGCCTGAACAGGTTTTTTTACTTCC 5792

Db 36 CAATAAATGAAGCCTGAACAGGTTTTTTTACTTCC 1

RESULT 7

US-09-918-995-7316
; Sequence 7316, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7316
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-7316

Query Match 5.9%; Score 389.4; DB 10; Length 391;
Best Local Similarity 99.7%; Pred. No. 1.3e-70;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4639 CTCACCTGAAACTCTAAAGAGTAATGGGATGAATAAAATGGAGTTCAGAGTTTCCATTG 4698
DB 1 CTCACCTGAAACTCTAAAGAGTAATGGGATGAATAAAATGGAGTTCAGAGTTTCCATTG 60

QY 4699 AACTGATATGGTGAATTTCAAGGAGGATCTGAAGGCCAACTTCTGCTCAGCATTTATC 4759
DB 61 AACTGATATGGTGAATTTCAAGGAGGATCTGAAGGCCAACTTCTGCTCAGCATTTATC 120

QY 4759 TAAATGATCTGATAGTCTGATACCTGTCATCCATGGTGGGACCTCCCACTCTAGTT 4818
DB 121 TAAATGATCTGATAGTCTGATACCTGTCATCCATGGTGGGACCTCCCACTCTAGTT 180

QY 4819 TACCATTAGAAATAGATTTAGTGTGTTTTCATTATAGAACATCTTTTGTAGTGAAGAATG 4878
DB 181 TACCATTAGAAATAGATTTAGTGTGTTTTCATTATAGAACATCTTTTGTAGTGAAGAATG 240

QY 4879 TGCCATATTACATATGCAACCTAATTTGTTAACTAACTCCAGCACTTAAGCTGAAT 4938
DB 241 TGCCATATTACATATGCAACCTAATTTGTTAACTAACTCCAGCACTTAAGCTGAAT 300

QY 4939 GCCACAAACACTAAAGATTAATATGTCGTGATTTTGAACACATAGCTTTGCTCTTT 4998
DB 301 GCCACAAACACTAAAGATTAATATGTCGTGATTTTGAACACATAGCTTTGCTCTTT 360

QY 4999 AGCAGGAATGATCTTTCAATCATTAGCA 5029
DB 361 AGCAGGAATGATCTTTCAATCATTAGCA 391

RESULT 8

US-10-085-783A-29775
; Sequence 29775, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29775
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t

US-10-085-783A-29775

Query Match 4.5%; Score 300.4; DB 13; Length 420;
Best Local Similarity 94.7%; Pred. No. 4.1e-52;
Matches 321; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 185 ATATCTCTCCAGGATGTTCTCAAGGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 244
DB 82 AAGCTCCGCGGGGCTGTAGGGGCGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 141

QY 245 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGGACAAACTCCTTGATGATTT 304

DB 142 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGGACAAACTCCTTGATGATTT 201

QY 305 TGAACAGAACCCAGATGAACAAGATTATCTCGCAGATGTACAAAGTGCATATGATTTCTAA 364

DB 202 TGAACAGAACCCAGATGAACAAGATTATCTCGCAGATGTACAAAGTGCATATGATTTCTAA 261

QY 365 CCAGTGTCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATTGCTCCCAAAGA 424

DB 262 CCAGTGTCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATTGCTCCCAAAGA 321

QY 425 CCAAGATGGCTTAATA-GTTGTGCTCTCATGAAACAGCTATGGAACAAATGAGATT 483

DB 322 CCAAGATGGCTTAATAAGTTGTTGCTCTCATGAAACAGCTATGGAACAAATGAGATT 381

QY 484 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATAC 522

DB 382 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATNC 420

RESULT 9

US-10-242-535A-29775
; Sequence 29775, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29775
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (419)..(419)
; OTHER INFORMATION: n is a, c, g, or t

US-10-242-535A-29775

Query Match 4.5%; Score 300.4; DB 16; Length 420;
Best Local Similarity 94.7%; Pred. No. 4.1e-52;
Matches 321; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 185 ATATCTCTCCAGGATGTTCTCAAGGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 244
DB 82 AAGCTCCGCGGGGCTGTAGGGGCGCATACAAAGATTAATAATCTGAATAAGTCTCAGG 141

QY 245 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGGACAAACTCCTTGATGATTT 304

DB 142 TAGGATGGACAGTTATTTAAAGCAGCTGTCAAGTGTGGACAAACTCCTTGATGATTT 201

QY 305 TGAACAGAACCCAGATGAACAAGATTATCTCGCAGATGTACAAAGTGCATATGATTTCTAA 364

DB 202 TGAACAGAACCCAGATGAACAAGATTATCTCGCAGATGTACAAAGTGCATATGATTTCTAA 261

QY 365 CCAGTGTCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATTGCTCCCAAAGA 424

DB 262 CCAGTGTCTCAGTTTCTTCAGAGTTGGCTTCTCCACAGCGAACTTCATTGCTCCCAAAGA 321

QY 425 CCAAGATGGCTTAATA-GTTGTGCTCTCATGAAACAGCTATGGAACAAATGAGATT 483

DB 322 CCAAGATGGCTTAATAAGTTGTTGCTCTCATGAAACAGCTATGGAACAAATGAGATT 381

Qy 484 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATAC 522
Db 382 CCCTGAATGAAAAACACTCAAGGAGCTTACTTCTATNC 420

RESULT 10

US-10-106-698-2133
; Sequence 2133, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA00591
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 2133
; LENGTH: 317
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (260)..(260)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (268)..(268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (293)..(293)
; OTHER INFORMATION: n equals a,t,g, or c
; US-10-106-698-2133

Query Match 4.3%; Score 286; DB 15; Length 317;
Best Local Similarity 95.6%; Pred. No. 3.4e-49;
Matches 303; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
Qy 1609 GAAGATAGATCTCTGACAGACAGTAATCAGAGTGTCTTTGGATGGTGGTACACCA 1668
Db 2 GAAGATAGATCTCTGACAGACAGTAATCAGAGTGTCTTTGGATGGTGGTACACCA 61
Qy 1669 GTTCTACAGTGTAGAACTCTCAAGAGGGCTTTCTGGCACTCATGCTCCAGAGCTTCTG 1728
Db 62 GTTCTACAGTGTAGAACTCTCAAGAGGGCTTTCTGGCACTCATGCTCCAGAGCTTCTG 121
Qy 1729 ATTGTTGTGAAGTGTATTAATCTTTTCAAGCAATGATGATGGGCAAGACTTAG 1788
Db 122 ATTGTTGTGAAGTGTATTAATCTTTTCAAGCAATGATGATGGGCAAGACTTAG 181
Qy 1789 ATTACTTTAATTAATGATGAGCGGCAAAAGTGGCCCACTAATTAATGATCTGAAGCTG 1848
Db 182 ATTACTTTAATTAATGATGAGCGGCAAAAGTGGCCCACTAATTAATGATCTGAAGCTG 240
Qy 1849 ATGCCCTTCTGACAGACAGTATCTTCAGACCCTAATCAATAAGTCTTTTGAAGAAATG 1908
Db 241 ATGCCCTTCTGACAGACAGTATCTTCAGACCCTAATCAATAAGTCTTTTGAAGAAATG 300
Qy 1909 TAAATGACTCTAAATCG 1925
Db 301 TAAATGACTCTAAATCG 317

RESULT 11

US-09-960-352-13367
; Sequence 13367, Application US/09960352
; Patent No. US20020137139A1

; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13367
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 57-LIB3058-036-Q1-K1-G2
US-09-960-352-13367

Query Match 3.0%; Score 199.6; DB 9; Length 393;
Best Local Similarity 81.4%; Pred. No. 3.4e-31;
Matches 311; Conservative 0; Mismatches 49; Indels 22; Gaps 6;
Qy 5111 TTTCTTTTAAAGAAATTTATAGCATTTACTGTGTATTAAATGCT-AAGCCAAAGTATCT 5169
Db 22 TGTCTTTATAGAAATTTTGTAGCATTTACTGTGTGTAAATGCTAAAGCCAAAGTATCT 81
Qy 5170 GCATTAGGTATACCTCTTTATGCCAATAATGATTTTATGAAGGCTCTTTTCAGATGA 5229
Db 82 GCATTCTTTTATGCCAAGAACGATTTTAGTGAAGGCTCTTTTCAGATGA 129
Qy 5230 ACCTATGAGGAATATCTGCTTTGTGTATATGCCAGTTAGAACTACTGTTTCTAAAGT 5289
Db 130 ACCTATGAGGAATATCTGCTTTGTGTATATGCCAGTTAGAACTACTGTTTCTAAAGT 189
Qy 5290 CTG-TCAAAATGTATTTCAGTGGCACAACCAAGTTTGGAGCTCTTGAAGCTT-AT 5343
Db 190 CTGTTAAATGTATTTCATGTCAC--AGACCAGTTTTCAGATCTTACACTATATGAT 247
Qy 5344 AATCTTTGATTAATAACTGA-TAATCTTTTGTATTAATGGAGTGAGACCTACCTCCAT 5402
Db 248 AATCTTTGATTAATAAGAGATTAATGATTAATTTGTTTAAATGGAGTGAGACCTACCTCCAT 307
Qy 5403 AATTAGATAAATCTTTTGGATTATAATCAGAAATTTGCCTTTTTCTTCTCAAAATAT 5462
Db 308 GATTAGATAAAGATTTATTTGCAATTAATAATCAGAAATTTGCCTTTTTCTTCTCAAAATAT 367
Qy 5463 TACATATGATGATTAATATATAT 5484
Db 368 AATTAAATATATATATATATATATAT 389

RESULT 12

US-10-040-739-971
; Sequence 971, Application US/10040739
; Publication No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaValle, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13497
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018764.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; US-09-864-761-13497

Query Match          2.6%; Score 172; DB 9; Length 556;
Best Local Similarity 100.0%; Pred. No. 2.3e-25;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4540 AGCTGTTCTACTTCTTAAGACACAGGATTTATCTATTTTATCAACTTCTTATCAGTTTG 4599
DB 208 AGCTGTTCTACTTCTTAAGACACAGGATTTATCTATTTTATCAACTTCTTATCAGTTTG 267
QY 4600 CAAAAGAAATAGCCATGCTTGTAGTGTGCGCTGTGCGCTCACCTGAAACACTCTAAAAA 4659
DB 268 CAAAAGAAATAGCCATGCTTGTAGTGTGCGCTGTGCGCTCACCTGAAACACTCTAAAAA 327
QY 4660 GTAATGGATGAATAAATTTGGATCTCAGATTTCCATTCACACTGATATGGT 4711
DB 328 GTAATGGATGAATAAATTTGGATCTCAGATTTCCATTCACACTGATATGGT 379

RESULT 14
US-09-864-761-30043
; Sequence 30043, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30043
; LENGTH: 168
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018764.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: g17662047, EVALUATE 5.00e-90
; OTHER INFORMATION: SWISSPROT HIT: Q05071, EVALUATE 8.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE883109.1, EVALUATE 7.00e-90
; US-09-864-761-30043

Query Match      2.5%; Score 168; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 7.6e-25;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4542 GTGTTCTACTTTCTAAAGGACGAGGATTTATCTATTTTATCAACTTCTTTATCAGTTTGCA 4601
Db      |||||||
1 GTGTTCTACTTTCTAAAGGACGAGGATTTATCTATTTTATCAACTTCTTTATCAGTTTGCA 60

QY 4602 AAAGAAATAGCCATGGCTTGTAGTCTGCGTGTGCGCTCCCTCAGTGAAGAACTCTAAAGAGT 4661
Db      |||||||
61 AAAGAAATAGCCATGGCTTGTAGTCTGCGTGTGCGCTCCCTCAGTGAAGAACTCTAAAGAGT 120

QY 4662 AATGGGATGAATAAATTGGAGCTCAGAGTTCCATTCAGACTGATATG 4709
Db      |||||||
121 AATGGGATGAATAAATTGGAGCTCAGAGTTCCATTCAGACTGATATG 168

RESULT 15
US-09-880-107-5/c
; Sequence 5, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
```

```
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA001902
; NAME/KEY: unsure
; LOCATION: (1)..(387)
; OTHER INFORMATION: n = a or c or g or t
; US-09-880-107-5

Query Match      1.9%; Score 124; DB 9; Length 387;
Best Local Similarity 96.9%; Pred. No. 1.7e-15;
Matches 157; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 6474 GTATGTATTCTG-CCATGTAAGTAATT-GAACAAGTCTTAAATAA-CCAAATGGTAGAGG 6530
Db      |||||||
381 GTATGTATTCTGCCCATGTAGTAATTTGGAACAGTCTTAAATAACCCCAATGGTAGAGG 322

QY 6531 GCTGTTCCATGATGGGACAGCTTTGGATTGTTTTCATAAAATCTCTACATTCATATAAAA 6590
Db      |||||||
321 GCTGTTCCATGATGGGACAGCTTTGGATTGTTTTCATAAAATCTCTACATTCATATAAAA 262

QY 6591 ATTGGAAATTATGTGCTGAGTTTGGAGGACACATTTTGAAGT 6632
Db      |||||||
261 ATTGGAAATTATGTGCTGAGTTTGGAGGACACATTTTGAAGT 220

Search completed: May 18, 2004, 14:34:39
Job time : 1714.6 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:54:15 ; Search time 12011.2 Seconds
(without alignments)
17461.734 Million cell updates/sec

Title: US-09-744-167-1

Perfect score: 4839

Sequence: 1 gcatatgtaatcagcaggac.....atcaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: , 5940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBankl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_hfgo_hum.*
- 40: em_hfgo_mus.*
- 41: em_hfgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4839	100.0	4839	9	AF104304	AF104304 Homo sapi
2	4679.8	96.7	4769	9	BC032680	BC032680 Homo sapi
3	4309.2	89.1	4565	9	AF130419	AF130419 Homo sapi
4	2397.6	49.5	2625	9	AF130420	AF130420 Homo sapi
5	2110.4	43.6	13799	9	AC105754	AC105754 Homo sapi
6	2110.4	43.6	150129	2	AC026909	AC026909 Homo sapi
7	1977	40.9	1977	6	AX809313	AX809313 Sequence
8	1552.2	32.1	89373	9	AL591398	AL591398 Human DNA
9	1545.8	31.9	4823	5	AF104305	AF104305 Xenopus l
C 10	1421.6	29.4	85802	10	AL671909	AL671909 Mouse DNA
C 11	1399.2	28.9	258787	2	AC095988	AC095988 Rattus no
C 12	1265.6	26.2	201330	2	AC118106	AC118106 Rattus no
13	898.8	18.6	1237	10	BC046341	BC046341 Mus muscu
14	875.4	18.1	164706	9	AC068614	AC068614 Homo sapi
15	740.8	15.3	1144	10	BC038255	BC038255 Mus muscu
16	699.4	14.5	888	6	AX677329	AX677329 Sequence
17	658	13.6	887	10	BC058416	BC058416 Mus muscu
18	576	11.9	576	6	AX677675	AX677675 Sequence
19	575.4	11.9	107627	9	AL513218	AL513218 Human DNA
C 20	575.4	11.9	123170	2	AL138792	AL138792 Homo sapi
C 21	575.2	11.9	192533	9	AC055876	AC055876 Homo sapi
C 22	564.8	11.7	151800	9	AC138649	AC138649 Homo sapi
C 23	564.6	11.7	178257	2	AC024350	AC024350 Homo sapi
24	564.4	11.7	3793	9	AK127003	AK127003 Homo sapi
25	562.2	11.6	6280	9	HSN805726	BS37424 Homo sapi
26	560.6	11.6	6632	9	AB002303	AB002303 Human mRN
27	559	11.6	5841	9	HSN804398	AL833057 Homo sapi
28	558.2	11.5	180741	9	AC091565	AC091565 Homo sapi
29	556.2	11.5	4620	9	AF434817	AF434817 Homo sapi
30	551.4	11.4	5438	10	AK122241	AK122241 Mus muscu
31	551.4	11.4	5480	10	BC042669	BC042669 Mus muscu
C 32	539	11.1	228470	2	AC136776	AC136776 Homo sapi
C 33	340	7.0	179370	10	AL626783	AL626783 Mouse DNA
C 34	339.6	7.0	5158	9	HSN809324	AX918602 Sequence
35	318	6.6	332	6	AX918602	BD054135 Sequence
36	318	6.6	332	6	BD054135	AC127085 Rattus no
37	313.4	6.5	193289	2	AC127065	AF239997 Drosophill
38	245.8	5.1	4598	3	AF239997	AJ310804 Drosophill
39	245.8	5.1	4651	3	DME310804	AY051821 Drosophill
40	244.2	5.0	4660	3	AY051821	AC019884 Drosophill
C 41	202	4.2	54186	2	AC019884	AC004564 Drosophill
C 42	202	4.2	67833	3	AC004564	AC093307 Drosophill
43	202	4.2	179139	3	AC093307	AE003454 Drosophill
44	202	4.2	313634	3	AE003454	BC032227 Homo sapi
45	190	3.9	4499	9	BC032227	

ALIGNMENTS

RESULT 1

AF104304 4839 bp mRNA linear PRI 03-JAN-1999

LOCUS AF104304 Homo sapiens Smad anchor for receptor activation (SARA) mRNA,

DEFINITION AF104304 complete cds.

ACCESSION AF104304

VERSION AF104304.1 GI:4092766

KEYWORDS SARA, a-FIVE domain protein that recruits Smad2 to the TGFbeta

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 4839)

AUTHORS Tsukazaki, T., Chiang, T.A., Davidson, A.F., Attisano, L. and Wrana, J.L.

TIME SARA, a-FIVE domain protein that recruits Smad2 to the TGFbeta

3541	Db	ATCAAAATTC	CCAGCAACAGATACAATGAGATGATGAAAGCCATGAACAAGTCCAATGAG	3560
3601	Qy	CATGTCCTGG	CAGGAGGTGCTGCTTCAATGAAAGGCAGAGCTCTCATCTTGTGTGTGTA	3660
3601	Db	CATGTCCTGG	CAGGAGGTGCTGCTTCAATGAAAGGCAGAGCTCTCATCTTGTGTGTGTA	3660
3661	Qy	CAGAAATGATG	TGGAATCTATCAGACCCAGGCTATCAGTATTCAATCAGCCAGCAAAA	3720
3661	Db	CAGAAATGATG	TGGAATCTATCAGACCCAGGCTATCAGTATTCAATCAGCCAGCAAAA	3720
3721	Qy	GTGACTGGT	GCCAGTTCTTTGTGTTCAGTGGGCGCTCTGAAATCCTCTTCTGGATACCTT	3780
3721	Db	GTGACTGGT	GCCAGTTCTTTGTGTTCAGTGGGCGCTCTGAAATCCTCTTCTGGATACCTT	3780
3781	Qy	GCCAAGTCC	AGTATTTGTGGAAGATGGTGTTATGGTCCAGATTA	3840
3781	Db	GCCAAGTCC	AGTATTTGTGGAAGATGGTGTTATGGTCCAGATTA	3840
3841	Qy	TCCTTGAGG	CAGCAGCTCGCAGAGATGAAGAGCTTCACCATCACTGTGGGAAGCGGCAC	3900
3841	Db	TCCTTGAGG	CAGCAGCTCGCAGAGATGAAGAGCTTCACCATCACTGTGGGAAGCGGCAC	3900
3901	Qy	GCAGAGAA	CCCCCAGGAGCAATCCACATCCAGTGGGTGGATGATCAACAAGACGTTAGC	3960
3901	Db	GCAGAGAA	CCCCCAGGAGCAATCCACATCCAGTGGGTGGATGATCAACAAGACGTTAGC	3960
3961	Qy	AAGGGTGCT	AGTACTATAGATGGGAAGTCCATGGAGACTATACAAATGTCAAGATA	4020
3961	Db	AAGGGTGCT	AGTACTATAGATGGGAAGTCCATGGAGACTATACAAATGTCAAGATA	4020
4021	Qy	TTCCATGGA	TACAAATATAAAGCAAAATGGAAAAAGTAAATCAGATGGAGCAGAGGTGTTTTTC	4080
4021	Db	TTCCATGGA	TACAAATATAAAGCAAAATGGAAAAAGTAAATCAGATGGAGCAGAGGTGTTTTTC	4080
4081	Qy	CTAGAAAA	CGATGACACAAATTTGCTCAGTGTATCTCTGCAGATCACAGTAGATTGACT	4140
4081	Db	CTAGAAAA	CGATGACACAAATTTGCTCAGTGTATCTCTGCAGATCACAGTAGATTGACT	4140
4141	Qy	GAGCATGTT	GCCAAAGCTTTTTCCTTGCTCTCTGCTCCTCACTGAAACTTCTCTGAAGGAA	4200
4141	Db	GAGCATGTT	GCCAAAGCTTTTTCCTTGCTCTCTGCTCCTCACTGAAACTTCTCTGAAGGAA	4200
4201	Qy	GATGGAAT	ACCAAACTGGGACTACGTGTGACACTTTGACTCAGATCAGGTGGCTATCAA	4260
4201	Db	GATGGAAT	ACCAAACTGGGACTACGTGTGACACTTTGACTCAGATCAGGTGGCTATCAA	4260
4261	Qy	GCAGGAGCA	ATGGCCAGCCCTTCCTCGACGTACATGAATGATCTGGATAGCGCCTTG	4320
4261	Db	GCAGGAGCA	ATGGCCAGCCCTTCCTCGACGTACATGAATGATCTGGATAGCGCCTTG	4320
4321	Qy	GTGCCGGT	GTATCCATGGAGGGGCTGCCAGCTTAGTGGGGCCCGCTTGTCTATGGAACCTC	4380
4321	Db	GTGCCGGT	GTATCCATGGAGGGGCTGCCAGCTTAGTGGGGCCCGCTTGTCTATGGAACCTC	4380
4381	Qy	ATCTTTTATA	TCTTGAAACATCGTATAAAACAGAGAAGACTTCATTTTTTCTGTTCAG	4440
4381	Db	ATCTTTTATA	TCTTGAAACATCGTATAAAACAGAGAAGACTTCATTTTTTCTGTTCAG	4440
4441	Qy	ACTTGTGTA	CCAGCAGTATACCCAAATCAITTTGACATTTTAAACTCGGAAGATTAAAGCT	4500
4441	Db	ACTTGTGTA	CCAGCAGTATACCCAAATCAITTTGACATTTTAAACTCGGAAGATTAAAGCT	4500
4501	Qy	TTTGTTTAA	CACTATTAAATGGGGTGGGAATAGGGTGGAGTGGGGGTTTTGGGAGACGGGT	4560
4501	Db	TTTGTTTAA	CACTATTAAATGGGGTGGGAATAGGGTGGAGTGGGGGTTTTGGGAGACGGGT	4560
4561	Qy	GGGAAAGGG	TGGTGTGGGGGACCGATGTTCCATTAATTTCTAAGTCTCTATGCAATGTCCA	4620
4561	Db	GGGAAAGGG	TGGTGTGGGGGACCGATGTTCCATTAATTTCTAAGTCTCTATGCAATGTCCA	4620
4621	Qy	CCAAGAAGAT	CTGGGAGCTTCTGTTCCTGCACAAACAGTTATGCTATCTTCAGACTAAT	4680
4621	Db	CCAAGAAGAT	CTGGGAGCTTCTGTTCCTGCACAAACAGTTATGCTATCTTCAGACTAAT	4680

Qy	4681	CCCTTCTGTTACTGTTTAGACAAGAATTCGGCTCTCTCTCTCAAGATTACTTAATCGTCA	4740
Db	4681	CCCTTCTGTTACTGTTTAGACAAGAATTCGGCTCTCTCTCTCAAGATTACTTAATCGTCA	4740
Qy	4741	TGTGCTCAGAAATGCTCAAATGGTGTAACACCATCACCAAGGGTGGCATGGAGGGCAGAG	4800
Db	4741	TGTGCTCAGAAATGCTCAAATGGTGTAACACCATCACCAAGGGTGGCATGGAGGGCAGAG	4800
Qy	4801	GGGAATAAATAATATAAGGCATCAAAAAAATTTTTTTTTT	4939
Db	4801	GGGAATAAATAATATAAGGCATCAAAAAAATTTTTTTTTT	4939

RESULT 2	BC032680	4769 bp	mRNA	linear	PRI 08-OCT-2003
LOCUS	BC032680	Homo sapiens MAD, mothers against decapentaplegic homolog (Drosophila) interacting protein, receptor activation anchor, mRNA (cDNA clone MGC:45259 IMAGE:558517), complete cds.			
ACCESSION	BC032680	1	GI:21618607		
KEYWORDS	MGC:				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 4769)				
	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Ahtshul,S.F., Zeebor,B., Buetow,K.H., Schafer,C.F., Bhat,N.K., Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P., Dietchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Staptenko,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Utsid,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Millaby,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Wolkman,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Viallon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahay,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butcherfield,Y.S., Krzywicki,M.I., Skalska,U., Smalls,D.E., Scherch,A., Schein,J.E., Jones,S.J., and Maizra,M.A.				
TITLE	Generation and initial analysis of more than 15,700 full-length human and mouse cDNA sequences				
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16895-16903 (2002)				
PUBMED	22388257				
REFERENCE	12477932				
AUTHORS	2 (bases 1 to 4769)				
TITLE	Strausberg,R.				
JOURNAL	Direct Submission				
REMARK COMMENT	Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
	NIH-MGC Project URL: http://mgc.nci.nih.gov				
	Contact: MGC help desk				
	Email: gcgaps@mail.nih.gov				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				
	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;				
	Web site: http://www.nisc.nih.gov				
	Contact: nisc.mgc@grni.nih.gov				
	Akhtar,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,				

Maduro, Q.L., Masiello, C., Maskeri, B., Mastriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
Series: IRAK Plate: 69 Row: p Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4759059.

FEATURES

source

1..4769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:45259 IMAGE:5585817"
/tissue_type="Testis, embryonal carcinoma"
/clone_lib="NIH MGC_92"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..4769
/gene="FIVE"
/note="synonyms: SARA, ZFYVE9"
/db_xref="LocusID:9372"
/db_xref="MIM:603755"
12..4289

gene

CDS

/product="MADHIP protein"
/protein_id="AAH32680.1"
/db_xref="GI:21618608"
/db_xref="LocusID:9372"
/translation="MENYFOAEAYNKVLDFEONEDETYSSTLLDTKWKILDPDS
HRLSFPILASVNSAESQPOLKVFSLAHAPLUTTEEDHCANGQDCNLPRIAT
MWIDENVAEDQLIKRNYSDWDCSAVEGKCGNLAELPDEKKNLVAVMENDCKR
TLQNDQCNNSQMDAPSCDLNENRQTFQFSINESKKNMSEKQMDPLNR
KPTGRSVNHLCTPSSLASVCSQSDKDDSGISDRDPSMAITSLTVDVSISSQGT
GCPAVKQENYIPEDLTGKISSPRTDLGSPNSFMSHSEGLMKRPAEBSFTEESLR
SGPLLLKPMQNGSNNDCRSCDLVNEVRADENEGYEHEETLGTTEFLNMTET
PSESQDMNKLTKLNEMDSQVNEKEKELQISOPEDTNGDSGGQVGLADGLDK
GTCISEEDCFSTVIDTPAAYLSNGDSIGWDPGVFVVKILSKESVSVEEREI
EESKESYNIYEGNEATTEGSLNLSNDGLMKNYLHNFQSPVSLVQSSPKV
ASLPSISVFPFGARPKFNLKQIPKPLDHLQNDPANSNNNTKNDILGAKLG
ENSNATVCSPLGNSVNDTNGHELYEAEISIRCLALAPSPNDLRAGFGISA
RKPTTLGEVAPVNDQAPNCKCEARFTTKRHHCRACGKVCSCSCLKKLL
YMDSEARVCVICHSLVMAQAWENMSASSOSPNPNPAEYCSIIPLQOASGAL
SSPPTVNPVGVVLKHPGAEVAQPREQRRVWEADGILPNGEVADAALKTWNGTSSAGT
LAVSHDPKPTVTPSPFAETDILCFSGSITQVGSFVSANMLIPEDGLPILISTGVK
GDYAVEKPSQISVYQLEDGDPPLVFLVNLNLSMVKIYVYVNRKWCFTTNGMHA
VGQSEIVLLQCLPDEKLPKDIENHFVQLYRDALAGNVSNLGHFSFOSFLGSKKH
GGFLYVSTYQSLQDLVLPYPLGILIKWETPWAKVPIRLMLRLGAEYRLYPCP
LPSVFRKPLFGTGTHTIMNLLAFRNVQVTLPVVQGLVDMVRKTSIKIPSNRYNE
MKAMKNSHVLGAGGACFEKADSHLVQVNDQNGYQTOAISHNOPRVKTGASPV
FSGALKSSGVYLAQSSIVEDGVWQITAEKNDSLRQALREMKOPTITCGKADAEERPE
HIHQVDDDDKNVSKGVSPIDGKSMETITNKKIFPHGEYKANGKVIWTEVFELEND
DQHNCLSDPADSHLDETHVAFCLCPHLKLEKEDGMTKLGLRVLDSDDQVGYQAG
SNGQPLPSQYNNLDSALVPVHGGACQLSEGPVYVMEILFIYLENIY"
2091..2282
/note="FIVE: Region: FIVE zinc finger. The FIVE zinc
finger is named after four proteins that it has been found
in: Fab1, YOTB/ZK632.12, Vac1, and EEA1. The FIVE finger
has been shown to bind two Zn++ ions. The FIVE finger has
eight potential zinc coordinating cysteine positions. Many
members of this family also include two histidines in a
motif R-HHC+XCG, where + represents a charged residue and
X any residue. We have included members which do not
conserve these histidine residues but are clearly related"
/db_xref="CDD:pfam01363"

misc_feature

96.7%; Score 4679.8; DB 9; Length 4769;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4692; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
ORIGIN

129 CCGATGGAGAAATTACTTCCAGCAGAGCTTACACCTGGGGCAAGGTGTAGATGAATT 188
9 CCGATGGAGAAATTACTTCCAGCAGAGCTTACACCTGGGGCAAGGTGTAGATGAATT 67
189 TGAACAAACAGAGATGAACAGATTTCTTCTACTTTATTTGGATACAAAGTGAATAAGAT 248
68 TGAACAAACAGAGATGAACAGATTTCTTCTACTTTATTTGGATACAAAGTGAATAAGAT 127
249 TCTAGATCCCCCTTCTCACCGGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATC 308
128 TCTAGATCCCCCTTCTCACCGGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATC 187
309 TGCAGTTTCTAATAGTGCACACCACTGAAGTCTTCTCCCTGGCTCAATTCAGCTCC 368
188 TGCAGTTTCTAATAGTGCACACCACTGAAGTCTTCTCCCTGGCTCAATTCAGCTCC 247
369 CTTGACCAAGAGAGAGGATCACTGTGCTAATGGACAGGACTGTAAATCTAAATCCAGA 428
248 CTTGACCAAGAGAGAGGATCACTGTGCTAATGGACAGGACTGTAAATCTAAATCCAGA 307
429 GATTGCCAATGTGGATGAATAATGCTTTCGACAGAGACCACTTAATTAAGAGAA 488
308 GATTGCCAATGTGGATGAATAATGCTTTCGACAGAGACCACTTAATTAAGAGAA 367
489 CTATAGTTGGGATGATCAATGCACTGCTTGAAGTGGGAGAGAGAAATGTGGAACCT 548
368 CTATAGTTGGGATGATCAATGCACTGCTTGAAGTGGGAGAGAGAAATGTGGAACCT 427
549 GGTCTGTCGCAGATGAGAGAAATGTTCTTGTGTAGCCGTGATGCATCAATCTGATGA 508
428 GGTCTGTCGCAGATGAGAGAAATGTTCTTGTGTAGCCGTGATGCATCAATCTGATGA 487
609 AAGGACATTAACAAAGGATTAACAGATTAATTAATTAATTAATTAATTAATTAATTA 668
488 AAGGACATTAACAAAGGATTAACAGATTAATTAATTAATTAATTAATTAATTAATTA 547
669 TGTCTTGTAGCTGTCTACTCGATAATGAACAGACAACTGATCAATTTAGTTTAGTAT 728
548 TGTCTTGTAGCTGTCTACTCGATAATGAACAGACAACTGATCAATTTAGTTTAGTAT 607
729 AATGAGTCCACTGAAAGAGATATGAATTCAGAGAACAAATGGATCCATTCGAATAGCC 788
608 AATGAGTCCACTGAAAGAGATATGAATTCAGAGAACAAATGGATCCATTCGAATAGCC 667
789 GAAACAGAGGGGAGATCTGTAAACCATCTGTGTCTTCTTCTTCTTCTTCTTCTTCT 848
668 GAAACAGAGGGGAGATCTGTAAACCATCTGTGTCTTCTTCTTCTTCTTCTTCTTCT 727
849 TGTCTGTTCCCTTCAATTAAGAGTACGGAAGTATAGGTAGAGACCCCTCCATGTC 908
728 TGTCTGTTCCCTTCAATTAAGAGTACGGAAGTATAGGTAGAGACCCCTCCATGTC 787
909 TGCAGTTACAAGTTTAAACGGTTGATTCAAGTAATCTCATCCAGGGAACAGATGGATGTC 968
788 TGCAGTTACAAGTTTAAACGGTTGATTCAAGTAATCTCATCCAGGGAACAGATGGATGTC 847
969 TGTCTGTTAAAAAGCAAGAGAACTATATACAGATGAGGACCTCATCTGGCAAAATCAGCTC 1028
848 TGTCTGTTAAAAAGCAAGAGAACTATATACAGATGAGGACCTCATCTGGCAAAATCAGCTC 907
1029 TCTTAGACAGATCTAGGAGTCCCAATTCCTTTTCCACATGAGTGAGGGGATTTTGTAT 1088
908 TCTTAGACAGATCTAGGAGTCCCAATTCCTTTTCCACATGAGTGAGGGGATTTTGTAT 967
1089 GAAAAAGAGCCACAGCAGAGAGAGACCACTCAAGAAATCCCTCCGCTCTGGTTTACCTTT 1148
968 GAAAAAGAGCCACAGCAGAGAGAGACCACTCAAGAAATCCCTCCGCTCTGGTTTACCTTT 1027
1149 GCTTCTCAACACAGACATGCTTAATCGGTCTGGAAGAAATATGACTGTCAAGGGTCTC 1208
1028 GCTTCTCAACACAGACATGCTTAATCGGTCTGGAAGAAATATGACTGTCAAGGGTCTC 1087
1209 AGATTGCTTGTGCTTAATGAAGTTAGGCTGTAGAAATGAAGTTTATGAACATGAAGA 1268

TLNDLQDNVNSO LMDARSCSLDNENRQDQFSPSINESTKDMSEKQMDPLNR
 PRTEGRS VNHLCPTSSDLSASVCPSPQKDGSGIRDPMSAITSIVDSVISQGD
 GCPAVKQENYIPDSDLTKISSRPTDLGSPNSFSHMSSEGLMKKEPAEESTTEESUR
 SGLPALLKPDMPVNGSRNDCSDCLVPNEVRADENEGYEHEETLGTTFELNMTBH
 FSESDMTNWKLTNENMDQOVNBEKEKFLQISQPDNTGDSGQCVGLADGLDLK
 GTCISESECDFSTVIDTPAANYLSNGCDSYGMDPGVSFVPKTLPSKEDSVTEKEI
 RESKSECVNLYEORNEATEGSGLLNSTGDLAKKNVYLNHFCVSQVPLGQSPKV
 ASLPSISVFPGRPKPQSNLKLQIPKPLSDHLQNDPPANSNNITKNNDILKAKUG
 ENSATVNSPGLSGLNINVDVTGHELSYEASISIRPCLALAPDSPDNDLRQFGISA
 RKPFTLLEGEVAPVDSQAPNCKCEARFTTKRHRHCRACGVKFCASCCSLKCKLL
 YMDRKEARVCVICHVLMNVAQPREQRVFPADGILPNGEVADAALKTNGTISAGTL
 AVSHDEKPVVTSPLPABTDICLFSGSI TOVSPVGSAMNLI PEDGLPPLILISGVGK
 DYAVEKPSQISVMQLEDGDPDLVFLVNLNLSMVKI VNYVNRKWCFTTKGMHVA
 GQSEBIVILLQCLPEKCLPDI FNFVQLYRDALAGNVSNLHSPFSQSLKSEKHEG
 GLVYVTSYQSLDVLVLPFLYFLGLQKRWETPWKVPFIRLMRLRGAETRLPCLP
 FSVPRKPLFQETGHTIMLLADFNRYQITLVPVQGVVDMEVNRKTSIKIPSNRYNEM
 MKANNKSHVLAGGACNEKADSHLVQVQNDGNYQOALSIHNPQKVTGASFVVF
 SGALKSSSYGLAKSIVBDGVVMVQITAEENDSLQALRDEGLHHHLWEGRGSTPGAH
 PHPVGG"

misc_feature 1475..1477

/gene="NSP"

/note="active site"

ORIGIN

Query Match 89.1%; Score 4309.2; DB 9; Length 4565;

Best Local Similarity 96.1%; Pred. No. 0;

Matches 4521; Conservative 0; Mismatches 3; Indels 180; Gaps 3;

QY	1	GCATCTGAATCAGCAGCACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCACCGAGA	60
DB	42	GCATCTGAATCAGCAGCACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCACCGAGA	101
QY	61	AGTCGTCTCTTATCACGTGTGAAGGGAAGAGTTTAAACAAGTCTCTTAAGTGGTG	120
DB	102	AGTCGTCTCTTATCACGTGTGAAGGGAAGAGTTTAAACAAGTCTCTTAAGTGGTG	161
QY	121	TTTCTCCAGGATGAGAAATTAATTCAGAGAGTTCACCTGGGACAGGTGTTA	180
DB	162	TTTCTCCAGGATGAGAAATTAATTCAGAGAGTTCACCTGGGACAGGTGTTA	220
QY	181	GATGAATTTGAACAAACGAGAGTGAACAGTCTTCTACTCTTATTGGATACAAAGTGG	240
DB	221	GATGAATTTGAACAAACGAGAGTGAACAGTCTTCTACTCTTATTGGATACAAAGTGG	280
QY	241	AATGAAGTTCTAGATCCCTCTTCCAGCGGTGTCAITTAACCTACTTTGGCAGGTGTG	300
DB	281	AATGAAGTTCTAGATCCCTCTTCCAGCGGTGTCAITTAACCTACTTTGGCAGGTGTG	340
QY	301	AATGAATCTGAGTTCTTAATGAGTCACACCAACTGAAGTCTTCTCCCTGGCTCAT	360
DB	341	AATGAATCTGAGTTCTTAATGAGTCACACCAACTGAAGTCTTCTCCCTGGCTCAT	400
QY	361	TCAGTCTCCCTGACACAGAGAGAGGATCACTGTGCTAATGACAGGAGTGTAACTCTA	420
DB	401	TCAGTCTCCCTGACACAGAGAGAGGATCACTGTGCTAATGACAGGAGTGTAACTCTA	460
QY	421	AATCCAGAGATGCCAATGTGATTTGATGAATAATGCTGTTCAGAAAGACCAAGTTAAT	480
DB	461	AATCCAGAGATGCCAATGTGATTTGATGAATAATGCTGTTCAGAAAGACCAAGTTAAT	520
QY	481	AAGAGAACTACTAGTTGGGATGATCAATGCACTGTGCTAATGAGTGGGAGAGAGAAATGT	540
DB	521	AAGAGAACTACTAGTTGGGATGATCAATGCACTGTGCTAATGAGTGGGAGAGAGAAATGT	580
QY	541	CGAAACCTGGCTGTGTCAGATGAGAGAAATGTTCTTGTGTGAGCCGTATGCAATAAC	600
DB	581	CGAAACCTGGCTGTGTCAGATGAGAGAAATGTTCTTGTGTGAGCCGTATGCAATAAC	640
QY	601	TGTGATAAAGGACATTACAAACGATTTCAGAGATTTGTAATTAATATAGTCAATCC	660
DB	641	TGTGATAAAGGACATTACAAACGATTTCAGAGATTTGTAATTAATATAGTCAATCC	700
QY	661	CTTATGAGATGCTTTTAGCTGTCTACATGATAATGAAACACAGACAACTGATCAATTTAGT	720

DB	701	CTTATGAGATGCTTTTAGCTGTTCCTGGGATAATGAAACAGACAACTGATCAATTTAGT	760
QY	721	TTTAGTATAAATGAGTCCACTGAAAAGATATGAATTCAGAGAAACAAATGGATCCATTTG	780
DB	761	TTTAGTATAAATGAGTCCACTGAAAAGATATGAATTCAGAGAAACAAATGGATCCATTTG	820
QY	781	AATAGACCGAAAACAGAGGGGAGATCTGTTAACCATCTCTGTCTCTACTTCATCTGATAGT	840
DB	821	AATAGACCGAAAACAGAGGGGAGATCTGTTAACCATCTCTGTCTCTACTTCATCTGATAGT	880
QY	841	CTAGCCAGTGTCTGTTTCCCTCTTCAAAATTAAGGATGACGGAGATATAGGTAGAGACCCC	900
DB	881	CTAGCCAGTGTCTGTTTCCCTCTTCAAAATTAAGGATGACGGAGATATAGGTAGAGACCCC	940
QY	901	TCCATGCTCGGATTAACAAGTTTAAAGGATGACGGAGATATAGGTAGAGACCCC	960
DB	941	TCCATGCTCGGATTAACAAGTTTAAAGGATGACGGAGATATAGGTAGAGACCCC	1000
QY	961	GGATGCTCTGCTTTAAAAGCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAA	1020
DB	1001	GGATGCTCTGCTTTAAAAGCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAA	1060
QY	1021	ATCAGCTCTCTTAGGACAGATCTAGGAGTCCAAATTCCTTTCCCATGAGTGAGGGG	1080
DB	1061	ATCAGCTCTCTTAGGACAGATCTAGGAGTCCAAATTCCTTTCCCATGAGTGAGGGG	1120
QY	1081	ATTTTGATGAAAAGAGCCAGGAGAGACCACTGAAGATCCCTCCGGTCTGGT	1140
DB	1121	ATTTTGATGAAAAGAGCCAGGAGAGACCACTGAAGATCCCTCCGGTCTGGT	1180
QY	1141	TTACCTTTCTCTCAAAACACAGACATGCTTAATGGGTCTGGAAGGAATAATGACTGTGAA	1200
DB	1181	TTACCTTTCTCTCAAAACACAGACATGCTTAATGGGTCTGGAAGGAATAATGACTGTGAA	1240
QY	1201	CGGTGTCAGATGCTTGTGCTTAATGAAGTTAGGCTGATGAAAATGAAGTTATGAA	1260
DB	1241	CGGTGTCAGATGCTTGTGCTTAATGAAGTTAGGCTGATGAAAATGAAGTTATGAA	1300
QY	1261	CATGAAGAACTCTTGGCACTACAGATTCCTTAATATGACAGAGCACTTCTCTGAATCT	1320
DB	1301	CATGAAGAACTCTTGGCACTACAGATTCCTTAATATGACAGAGCACTTCTCTGAATCT	1360
QY	1321	CAGGACATGCTAAATGGAAGTTGACTAAACTTAATGATGATGATGATGATGATGATGATG	1380
DB	1361	CAGGACATGCTAAATGGAAGTTGACTAAACTTAATGATGATGATGATGATGATGATGATG	1420
QY	1381	GAGAAAGGAAAGTTTCTACAGATTTAGTCACTGAGGACACTTAATGATGATGATGATG	1440
DB	1421	GAGAAAGGAAAGTTTCTACAGATTTAGTCACTGAGGACACTTAATGATGATGATGATG	1480
QY	1441	GGACAGTGTGTTGGATGGCAGATCGAGTCTAGATTTTAAAGGAACTTGCATTAAT	1500
DB	1481	GGACAGTGTGTTGGATGGCAGATCGAGTCTAGATTTTAAAGGAACTTGCATTAAT	1540
QY	1501	AGTGAAGATGATGATTTCTCCACTTATAGACACACAGCAGCAAAATTTCTATCAAT	1560
DB	1541	AGTGAAGATGATGATTTCTCCACTTATAGACACACAGCAGCAAAATTTCTATCAAT	1600
QY	1561	GTTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACTTTACCC	1620
DB	1601	GTTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACTTTACCC	1660
QY	1621	TCCAAAGAGATTCAGTAAACAGAGAAAAGAAATAGAGAAAGCACTGAGATGCTAC	1680
DB	1661	TCCAAAGAGATTCAGTAAACAGAGAAAAGAAATAGAGAAAGCACTGAGATGCTAC	1720
QY	1681	TCAAAATTTTATGAACAGAGAGAAATCAGGCCCACAGAGGAGTGGACTACTTTTAAAC	1740
DB	1721	TCAAAATTTTATGAACAGAGAGAAATCAGGCCCACAGAGGAGTGGACTACTTTTAAAC	1780
QY	1741	AGCACTGGTACCTTAATGAAGAAAATTTATTATTAATTTCTGTAGTCAAGTTCCATCA	1800

Db 1781 AGCACTGGTGACCTAATGAAGAAAAATTTATTCATTAATTTCTGTAGTCAAGTTCATCA 1840
Qy 1801 GTGCTTGGGCAATCTTCCCCCAAGGTAGTAGCAAGCTGCCATCTATFCACTGTTCTCTTTT 1860
Db 1841 GTGCTTGGGCAATCTTCCCCCAAGGTAGTAGCAAGCTGCCATCTATFCACTGTTCTCTTTT 1900
Qy 1861 GGTGGTGAAGCAAGCAAGCAAGCAAGCTTCTAATCTTAACTTCAAACTTCAAAAGCAATATCA 1920
Db 1901 GGTGGTGAAGCAAGCAAGCAAGCAAGCTTCTAATCTTAACTTCAAACTTCAAAAGCAATATCA 1960
Qy 1921 GACCAATTTACAAATGACTTTCTGCAAGCAAGTGGAAATTAATCTAATAAATAAATAATGAT 1980
Db 1961 GACCAATTTACAAATGACTTTCTGCAAGCAAGTGGAAATTAATCTAATAAATAAATAATGAT 2020
Qy 1981 ATTCTTGGGAAGCAAAATAGGGGAAGCACTCAGCAACCAATGATGATGATGATGATGATGATG 2040
Db 2021 ATTCTTGGGAAGCAAAATAGGGGAAGCACTCAGCAACCAATGATGATGATGATGATGATGATG 2080
Qy 2041 GGAACATCTCTAATGTCGATACAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATC 2100
Db 2081 GGAACATCTCTAATGTCGATACAAATGGGAACATTTAGAAAGTTATGAGGCTGAGATC 2140
Qy 2101 TCCACTAGACATGCTTGCATTAGCTCCAGATAGCCCAAGATATGATCTCAGAGCTGGT 2160
Db 2141 TCCACTAGACATGCTTGCATTAGCTCCAGATAGCCCAAGATATGATCTCAGAGCTGGT 2200
Qy 2161 CAGTTTGGAAATTTCTGCCAGAAAGCAATCACCAGCTGGGTGAGTGGTCCAGTATGG 2220
Db 2201 CAGTTTGGAAATTTCTGCCAGAAAGCAATCACCAGCTGGGTGAGTGGTCCAGTATGG 2260
Qy 2221 GTACCGGATCTCAGGCTCCAAATGTCATGAAATGTGAAGCAGGTTTACATTCACCAAA 2280
Db 2261 GTCACGATCTCAGGCTCCAAATGTCATGAAATGTGAAGCAGGTTTACATTCACCAAA 2320
Qy 2281 AGGAGGCATCTGAGAGCATGTGGGAAGTCTTCTGTGCTTCTGCTGAGGCTGAAA 2340
Db 2321 AGGAGGCATCTGAGAGCATGTGGGAAGTCTTCTGTGCTTCTGCTGAGGCTGAAA 2380
Qy 2341 TGTAACTGTTTATACATGACAGAAAGCAAGCTAGAGTGTGTGTAATCTGCAATTCAGTG 2400
Db 2381 TGTAACTGTTTATACATGACAGAAAGCAAGCTAGAGTGTGTGTAATCTGCAATTCAGTG 2440
Qy 2401 CTAATGATGCTCAAGCCTGGGAGAACATGATGAGTGTGCTTCAAGCCAGAGCCCTAACCT 2460
Db 2441 CTAATGAT----- 2449
Qy 2461 AACATCTGCTGAATCTGTTCTACTATCCCTCCCTTGCAGCAAGCTCAGGCCCTCAGGA 2520
Db 2481 ----- 2449
Qy 2521 GCTCTGAGCTCTCCACCTCCCACTGTGATGATGATGATGATGATGATGATGATGATGATG 2580
Db 2541 ----- 2449
Qy 2581 GCAGAGTGGCTCAGCCAGAGAGCAGAGGCGAGTTTGGTTGCTGATGGATCTTGCC 2640
Db 2621 -----GTGGCTCAGCCAGAGAGCAGAGGCGAGTTTGGTTGCTGATGGATCTTGCC 2503
Qy 2641 AATGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
Db 2681 AATGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2563
Qy 2701 CTGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2760
Db 2741 CTGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2623
Qy 2761 GATATTTCTGATTTCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2820
Db 2801 GATATTTCTGATTTCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2683
Qy 2821 AATCTTATTTCTGAGAGTGGCTTCTCCCTCCATTTCTCCTCCTCCTCCTCCTCCTCCTCCT 2880
Db 2861 AATCTTATTTCTGAGAGTGGCTTCTCCCTCCATTTCTCCTCCTCCTCCTCCTCCTCCTCCT 2743

Qy 2881 TATGCTGTGGAAAGAAACCACATCAGATTTTCAGTAATGCAGCAGTTGGAGGATGGTGGC 2940
Db 2744 TATGCTGTGGAAAGAAACCACATCAGATTTTCAGTAATGCAGCAGTTGGAGGATGGTGGC 2803
Qy 2941 CTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTGTAATGTTTAAATTTGTTTAAAT 3000
Db 2804 CTGACCCACTTGTATTTGTTTAAATGCAAAATTTGTTGTAATGTTTAAATTTGTTTAAAT 2863
Qy 3001 TATGTAACAGAAAGTGTGGTGTTCACAAAGGGAATGCATGCAAGTGGGTGAGTCT 3060
Db 2864 TATGTAACAGAAAGTGTGGTGTTCACAAAGGGAATGCATGCAAGTGGGTGAGTCT 2923
Qy 3061 GAGATGATCATTTCTTCTCAGTGTTCACGGATGAAAGTGTTCGCAAGGATATCTTT 3120
Db 2924 GAGATGATCATTTCTTCTCAGTGTTCACGGATGAAAGTGTTCGCAAGGATATCTTT 2983
Qy 3121 AATCACTTTGTGAGCTTTTATCGGATGCTTGGCAGGGAATGTGGTGAACATTTGGG 3180
Db 2984 AATCACTTTGTGAGCTTTTATCGGATGCTTGGCAGGGAATGTGGTGAACATTTGGG 3043
Qy 3181 CATTCCTTCTCAGTCAAAAGTTCCTTGGCAGTAAAGACATGTTGGATTTCTTATATG 3240
Db 3044 CATTCCTTCTCAGTCAAAAGTTCCTTGGCAGTAAAGACATGTTGGATTTCTTATATG 3103
Qy 3241 ACATCTACCTACCAAGTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTTGGG 3300
Db 3104 ACATCTACCTACCAAGTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTTGGG 3163
Qy 3301 ATTCTTATCCAGAAATGGGAACTCTTGGGCTAAAGTATTTCTTATCGTCTGATGTTG 3360
Db 3164 ATTCTTATCCAGAAATGGGAACTCTTGGGCTAAAGTATTTCTTATCGTCTGATGTTG 3223
Qy 3361 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTGAGATTTGGAAG 3420
Db 3224 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTGAGATTTGGAAG 3283
Qy 3421 CCATTTTGGAGAGACGGGGCATCCATCATGAATCTTCTTGGCAGCTTCAGAAATTC 3480
Db 3284 CCATTTTGGAGAGACGGGGCATCCATCATGAATCTTCTTGGCAGCTTCAGAAATTC 3343
Qy 3481 CAGTATACCTTCCAGTGTTCAGGTTTCAAGGTTTGGTGTGATATGGAAGTTCGGAAGCTAGC 3540
Db 3344 CAGTATACCTTCCAGTGTTCAGGTTTCAAGGTTTGGTGTGATATGGAAGTTCGGAAGCTAGC 3403
Qy 3541 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAAGCCATGAACAGTCCAAATGAG 3600
Db 3404 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAAGCCATGAACAGTCCAAATGAG 3463
Qy 3601 CATGCTCTGAGAGGCTGCTGCTTCAATGAAAGGAGACTCTCATCTTGTGTGTA 3660
Db 3464 CATGCTCTGAGAGGCTGCTGCTTCAATGAAAGGAGACTCTCATCTTGTGTGTA 3523
Qy 3661 CAGAATGATGATGAAACTATCAGACCCAGGCTTATCAGTATTCACAATCAGCCAGAAAA 3720
Db 3524 CAGAATGATGATGAAACTATCAGACCCAGGCTTATCAGTATTCACAATCAGCCAGAAAA 3583
Qy 3721 GTCACTGGTCCAGTGTCTTGTGTTTCAAGTGGGCTCTGAAATCTCTTCTGTGATACCTT 3780
Db 3584 GTCACTGGTCCAGTGTCTTGTGTTTCAAGTGGGCTCTGAAATCTCTTCTGTGATACCTT 3643
Qy 3781 GCCAAGTCCAGTATTCGGAAGATGTTTATGTTCCAGATTAATGATGAGAAACATGGAT 3840
Db 3644 GCCAAGTCCAGTATTCGGAAGATGTTTATGTTCCAGATTAATGATGAGAAACATGGAT 3703
Qy 3841 TCCTTGGAGGAGGACCTGCGAGAGATGAAGGACTTCACCATCCTCTGTTGGGAAGGGGAC 3900
Db 3704 TCCTTGGAGGAGGACCT--GAGAGATGAAGGACTTCACCATCCTCTGTTGGGAAGGGGAC 3761
Qy 3901 GCGAGAGAACCCAGAGAGACATCCACATCCAGTGGGTGGATGATGAGAAAGCTTAGC 3960
Db 3762 GCGAGAGAACCCAGAGAGACATCCACATCCAGTGGGTGGATGATGAGAAAGCTTAGC 3821

Db 401 TCAGTCCCTGACCCAGAGGAGGATCACTGTGTAATGACAGGACTGTAATCTA 460
Qy 421 AATCCAGAGATGCCACAATGCGATGATGAATAATGCTGTGCCAGAGCCAGTAATTT 480
Db 461 AATCCAGAGATGCCACAATGCGATGATGAATAATGCTGTGCCAGAGCCAGTAATTT 520
Qy 481 AAGAGAACTATAGTTGGGATGATCAATGAGTGCCTGTTGAAGTGGGAGAGAGAAATGT 540
Db 521 AAGAGAACTATAGTTGGGATGATCAATGAGTGCCTGTTGAAGTGGGAGAGAGAAATGT 580
Qy 541 GGAAACCTGCTGTGTCGAGATGAGAGAAATGTTCTGTTGATGCGCTCATGCAATAC 600
Db 581 GGAAACCTGCTGTGTCGAGATGAGAGAAATGTTCTGTTGATGCGCTCATGCAATAC 640
Qy 601 TGTGATAAAGGACATTAACAAACGATTTACAGATTTGTAATTAATTAATTAATTAATTC 660
Db 641 TGTGATAAAGGACATTAACAAACGATTTACAGATTTGTAATTAATTAATTAATTAATTC 700
Qy 661 CTTATGATGCTTTTAGTCTGTCACTGGATTAATCAAAACAGACAAACTGATCAATTTAGT 720
Db 701 CTTATGATGCTTTTAGTCTGTCACTGGATTAATCAAAACAGACAAACTGATCAATTTAGT 760
Qy 721 TTTAGTATAATGATGCTCACTGAAAGATATGAATTCAGAGAAACAAATGGATCCATTG 780
Db 761 TTTAGTATAATGATGCTCACTGAAAGATATGAATTCAGAGAAACAAATGGATCCATTG 820
Qy 781 AATAGACGAAAAACAGAGGGGAGATCTGTAAACCATCTGTGCTCTACTTCATCTGATAGT 840
Db 821 AATAGACGAAAAACAGAGGGGAGATCTGTAAACCATCTGTGCTCTACTTCATCTGATAGT 880
Qy 841 CTAGCCAGTGTCTTCCCTTCAAAATTAAGAGATGACGGAATATAGTAGAGACCCC 900
Db 881 CTAGCCAGTGTCTTCCCTTCAAAATTAAGAGATGACGGAATATAGTAGAGACCCC 940
Qy 901 TCCATGCTCGATTAACAAGTTTAACGGTGAATTCAGTAACTCTCATCCAGGGAACAGAT 960
Db 941 TCCATGCTCGATTAACAAGTTTAACGGTGAATTCAGTAACTCTCATCCAGGGAACAGAT 1000
Qy 961 GGATGTCTGTGTTTAAAAAGCAAGAACTATATACAGATGAGGACCTCACTGGCAAA 1020
Db 1001 GGATGTCTGTGTTTAAAAAGCAAGAACTATATACAGATGAGGACCTCACTGGCAAA 1060
Qy 1021 ATCAGCTCTCTAGACAGATCTAGGAGTCCAAATCTTTTCCACATGAGTGAGGG 1080
Db 1061 ATCAGCTCTCTAGACAGATCTAGGAGTCCAAATCTTTTCCACATGAGTGAGGG 1120
Qy 1081 ATTTTGATGAAAAAGAGCCAGAGAGAGAGACCACTGAAGAATCCCTCCGCTCTGGT 1140
Db 1121 ATTTTGATGAAAAAGAGCCAGAGAGAGAGACCACTGAAGAATCCCTCCGCTCTGGT 1180
Qy 1141 TTACCTTTGCTTCTCAAAACAGACATGCTTAATGGGTCTGGAAGGAATATGATGTGAA 1200
Db 1181 TTACCTTTGCTTCTCAAAACAGACATGCTTAATGGGTCTGGAAGGAATATGATGTGAA 1240
Qy 1201 CGGTGTTCCAGATGCTTGTGCTTAATCAAGTATAGGCTGATGAAATGAAGGTTATGAA 1260
Db 1241 CGGTGTTCCAGATGCTTGTGCTTAATCAAGTATAGGCTGATGAAATGAAGGTTATGAA 1300
Qy 1261 CATGAAGAACTCTTGGCACTACAGATTCCTTAATGAAGAGATTTCTCTGAATCT 1320
Db 1301 CATGAAGAACTCTTGGCACTACAGATTCCTTAATGAAGAGATTTCTCTGAATCT 1360
Qy 1321 CAGGACATGACTAATTTGAGTTGACTAACTAATGATGATGATGATGATGATGATGATGAT 1380
Db 1361 CAGGACATGACTAATTTGAGTTGACTAACTAATGATGATGATGATGATGATGATGATGAT 1420
Qy 1381 GAAGAAAAAGGAAAAAGTTTCTACAGATTTAGTCCAGCTGAGGACCTAATTTGATGATGGA 1440
Db 1421 GAAGAAAAAGGAAAAAGTTTCTACAGATTTAGTCCAGCTGAGGACCTAATTTGATGATGGA 1480
Qy 1441 GGACAGTGTGTTGGATTTGGCAGATCCAGCTCTAGATTTTAAAGGACCTGATTTAGTGA 1500

RESULT 5
AC105754
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP4-814E15, complete sequence.
AC105754 AL122009
AC105754.3 GI:27228871
VERSION

Db 1481 GGACAGTGTGTTGGATTTGGCAGATGCGAGGTCTAGATTTTAAAGGAACTTGCAATTAGTCAA 1540
Qy 1501 AGTGAAGAATGTGATTTTCTCCACTGTTATAGACACACAGCAGCAAAATATCTATCTAAT 1560
Db 1541 AGTGAAGAATGTGATTTTCTCCACTGTTATAGACACACAGCAGCAAAATATCTATCTAAT 1600
Qy 1561 GGTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACCTTTACCC 1620
Db 1601 GGTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCCTAAAGACCTTTACCC 1660
Qy 1621 TCCAAAGAAAGATTCTAGTAAACAGAAAAAGAAATAGAGGAAAGCAAGTCAGAATGCTAC 1680
Db 1661 TCCAAAGAAAGATTCTAGTAAACAGAAAAAGAAATAGAGGAAAGCAAGTCAGAATGCTAC 1720
Qy 1681 TCCAAATATTATGAAACAGAGAGGAAATGAGGCCACAGAAAGGAGTGGACTACTTTTAAAC 1740
Db 1721 TCCAAATATTATGAAACAGAGAGGAAATGAGGCCACAGAAAGGAGTGGACTACTTTTAAAC 1780
Qy 1741 AGCACTGTGCTTAATGAAGAAAAATTTTACATAATTTCTGTAGTCAAGTTCCTATTT 1800
Db 1781 AGCACTGTGCTTAATGAAGAAAAATTTTACATAATTTCTGTAGTCAAGTTCCTATTT 1840
Qy 1801 GTGCTTTGGGCAATCTTCCCCCAAGGTAGTCAAGCCTGCGCATCTATCAGTGTTCCTTTT 1860
Db 1841 GTGCTTTGGGCAATCTTCCCCCAAGGTAGTCAAGCCTGCGCATCTATCAGTGTTCCTTTT 1900
Qy 1861 GGTGTGCAAGACCCAGCAACCTTCTAATCTTAACCTTCAAAATTCCAAAGCCATTATCA 1920
Db 1901 GGTGTGCAAGACCCAGCAACCTTCTAATCTTAACCTTCAAAATTCCAAAGCCATTATCA 1960
Qy 1921 GACCAATTTACAAATGACTTTCTCTGCAACAGTGGAAATTAATTAATAAATAAATGAT 1980
Db 1961 GACCAATTTACAAATGACTTTCTCTGCAACAGTGGAAATTAATTAATAAATAAATGAT 2020
Qy 1981 ATTCTTTGGAAAGCAAAATTAGGGAAACTCTCAGCAACCAATGTATGAGTCCATCTTTG 2040
Db 2021 ATTCTTTGGAAAGCAAAATTAGGGAAACTCTCAGCAACCAATGTATGAGTCCATCTTTG 2080
Qy 2041 GGAAACATCTCTAATGTGATACAAATGGGAAACATTTAGAAAGTTTATGAGGCTGAGATC 2100
Db 2081 GGAAACATCTCTAATGTGATACAAATGGGAAACATTTAGAAAGTTTATGAGGCTGAGATC 2140
Qy 2101 TCCACTAGACCATGCTTCCATTTAGCTCCAGATAGCCAGATGAATGATCTCAGAGCTGT 2160
Db 2141 TCCACTAGACCATGCTTCCATTTAGCTCCAGATAGCCAGATGAATGATCTCAGAGCTGT 2200
Qy 2161 CAGTTTGGAAATTTCTCCAGAAAGCCATTCAACGCTGGGTGAGTGGCTCCAGTATGG 2220
Db 2201 CAGTTTGGAAATTTCTCCAGAAAGCCATTCAACGCTGGGTGAGTGGCTCCAGTATGG 2260
Qy 2221 GTACCGGATTTCTCAGGCTCCAAATTTGCATGAAATGTGAAGCCAGGTTTACATTCACCAA 2280
Db 2261 GTACCGGATTTCTCAGGCTCCAAATTTGCATGAAATGTGAAGCCAGGTTTACATTCACCAA 2320
Qy 2281 AGAGAGCATCTGCAAGCATGTGGGAAGGTTTCTGTGCTTCTCTGCTGTAGCCTGAA 2340
Db 2321 AGAGAGCATCTGCAAGCATGTGGGAAGGTTTCTGTGCTTCTCTGCTGTAGCCTGAA 2380
Qy 2341 TGTAAACTGTTATACATGGAAGAGGAGTGTAGTGTGTAAATCTGCCATTCAGTGT 2400
Db 2381 TGTAAACTGTTATACATGGAAGAGGAGTGTAGTGTGTAAATCTGCCATTCAGTGT 2440
Qy 2401 CTAATGAATGCTCAAGCCTGGGAGAACATGAT 2432
Db 2441 CTAATGAATGCTCAAGCCTGGGAGAACATGAT 2472

HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 133799)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 133799)
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 REFERENCE 3 (bases 1 to 133799)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (26-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 REFERENCE 4 (bases 1 to 133799)
 AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 COMMENT On Dec 18, 2002 this sequence version replaced gi:18921345.

 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgctgs@u.washington.edu
 Drafting Center: SC

 Project Information
 Center project name: chr-1
 Center clone name: RP4-814E15 (sc0216)

 Summary Statistics
 Sequencing vector: plasmid; 56% of reads
 Sequencing vector: plasmid; 108752; 44% of reads
 Chemistry: Dye-terminator ET; 50% of reads
 Chemistry: Dye-terminator Big Dye; 50% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 133740 bases at least Q40
 Consensus quality: 133797 bases at least Q30
 Consensus quality: 133799 bases at least Q20
 Insert size: 133799; sum-of-contigs
 Quality coverage: 8.6x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': RP4-800M22 AL139156, 42347-bp overlap
 3': RP11-155018 AL513218, 2000-bp overlap

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

 This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

 ECORI BglII HindIII
 SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

 5810 5855 7542 7542 1289 1229
 2184 2142 5671 5692 449 <800
 8065 8190 5077 5096 512 <800
 486 <800 76 <800 2814 2825
 106 <800 3560 1247 1229
 4616 4708 414 <800 4695 4762
 14239 14077 1723 1693 3373 3356
 3925 3921 1216 1202 3274 3356
 983 984 2887 2916 3116 3125
 2594 2609 8059 7992 68 <800
 2871 2878 1854 1835 1886 1926
 2276 2254 4987 4925 3144 3125
 2711 2609 13600 12991 9189 9097
 4738 4708 2718 2784 1213 1229
 445 <800 2884 2916 1915 1926
 1531 1495 4547 4472 731 747
 3546 3538 467 <800 3148 3125
 3154 3144 900 919 3481 3356
 896 984 2356 2329 2570 2594
 802 884 134 <800 5185 5250
 11042 10836 4510 4472 868 865
 4907 4887 3070 3044 18109 18317
 6450 6524 2284 2329 12807 12493
 3079 3045 3751 3739 4972 4989
 532 <800 2586 2640 4206 4097
 773 777 1052 1063 374 <800
 1464 1495 1058 1063 1931 1926
 6249 6230 236 <800 11410 11255

[illegible]

1581 GCAAGACCCAGGTTCTTTTGTTCCTCAAGACTTTTACCTCCAAAGAGATTCAGTAAC 1640
 Db 61467 GCAAGACCCAGGTTCTTTTGTTCCTCAAGACTTTTACCTCCAAAGAGATTCAGTAAC 61526
 QY 1641 AGAAGAAAAGAAATAGAGAAAGCAAGTCAAGTCTACTCAATATTTATGAAACAGAG 1700
 Db 61527 AGAAGAAAAGAAATAGAGAAAGCAAGTCAAGTCTACTCAATATTTATGAAACAGAG 61586
 QY 1701 AGAATAGGCGCAGAGAGGAGTGGACTACTTTTAAACACACTGTCACCTTAATGAA 1760
 Db 61587 AGAATAGGCGCAGAGAGGAGTGGACTACTTTTAAACACACTGTCACCTTAATGAA 61646
 QY 1761 GAAAATATTTACATAAATTTCTGTAGTCAAGTTCATCAGTCTTGGGCAATCTCCCC 1820
 Db 61647 GAAAATATTTACATAAATTTCTGTAGTCAAGTTCATCAGTCTTGGGCAATCTCCCC 61706
 QY 1821 CAAGGTAGTCAAGAGCTCCCATCTATCAGTCTTCTTTGGTGTGCAAGACCAAGCA 1880
 Db 61707 CAAGGTAGTCAAGAGCTCCCATCTATCAGTCTTCTTTGGTGTGCAAGACCAAGCA 61766
 QY 1881 ACCTTCTAATCTAACTTCAATTCCTCAAGGCAATATCAGACCAATTTACAAAATGACTT 1940
 Db 61767 ACCTTCTAATCTAATCTCAATTCCTCAAGGCAATATCAGACCAATTTACAAAATGACTT 61826
 QY 1941 TCTGCAACACAGTGAATAATCTAATAAATAAATAAATGATATCTTTGGGAAAGCAAAAT 2000
 Db 61827 TCTGCAACACAGTGAATAATCTAATAAATAAATAAATGATATCTTTGGGAAAGCAAAAT 61886
 QY 2001 AGGGGAAACTCAGCAACCAATGTAGTGCAGTCCATCTTTGGGAAACATCTTAATGCGA 2060
 Db 61887 AGGGGAAACTCAGCAACCAATGTAGTGCAGTCCATCTTTGGGAAACATCTTAATGCGA 61946
 QY 2061 TACAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCATAGACCATGCTTGC 2120
 Db 61947 TACAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATCTCCATAGACCATGCTTGC 62006
 QY 2121 ATTAGCTCAGATAGCCAGATATGATCTAGAGCTGCTGAGTTTGAATTTCTGCGAG 2180
 Db 62007 ATTAGCTCAGATAGCCAGATATGATCTAGAGCTGCTGAGTTTGAATTTCTGCGAG 62066
 QY 2181 AAAGGCATTCACACGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTTCTAGGCTCC 2240
 Db 62067 AAAGGCATTCACACGCTGGGTGAGTGGCTCCAGTATGGGTACCGGATTTCTAGGCTCC 62126
 QY 2241 AAATTCATGAAATGTAAGCCAGTTTACATTCACAAAAGGAGGATCACTGCAGAGC 2300
 Db 62127 AAATTCATGAAATGTAAGCCAGTTTACATTCACAAAAGGAGGATCACTGCAGAGC 62186
 QY 2301 ATGTGGGAAGGT 2312
 Db 62187 ATGTGGGAAGGT 62198
 RESULT 6
 AC026909 150129 bp DNA linear HTG 19-APR-2000
 LOCUS Homo sapiens chromosome 1 clone RP11-174F10 map 1, WORKING DRAFT
 DEFINITION SEQUENCE, 27 unordered pieces.
 ACCESSION AC026909
 VERSION AC026909.2 GI:7596849
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 150129)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 1, clone RP11-174F10
 JOURNAL Unpublished
 2 (bases 1 to 150129)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy L., Boukhalter B., Brown A., Burkett, G.,
 Campopiano A., Castle A., Choepel Y., Colangelo M., Collins S.,
 Collymore A., Cooke P., DeArellano K., Dewar K., Diaz J.S.,
 Dodge S., Domino M., Doyle M., Ferreira P., FitzHugh W., Gage D.,
 Galand J., Gardyna S., Ginde S., Goyette M., Graham L.,
 Grand-Pierre N., Grant G., Hagos B., Heaford A., Horton L.,
 Klein J.C., Iliev I., Johnson R., Jones C., Kann L., Karatas A.,
 Kiew J., LaRoque K., Lamazares R., Landers T., Lehoczy J.,
 Levine R., Liu G., Locke K., MacDonald P., Marquis N.,
 McCarthy M., McEwan P., McGurk A., McKernan K., McPheeters R.,
 Meldrim J., Meneus L., Mihova T., Miranda C., Mienga V., Morrow J.,
 Murphy T., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,
 O'Neil D., Oliver T.M., Oliver J., Peterson K., Pierre N.,
 Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D.,
 Roy A., Santos R., Schauer S., Severy P., Spencer B.,
 Stange-thomann N., Stojanovic N., Subramanian A., Talamas J.,
 Tesfaye S., Theodore J., Tirrell A., Travers M., Trigilio J.,
 Vassiliev H., Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J.,
 Young G., Zainoun J., Zimmer A. and Zody M.
 Direct Submission
 Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 19, 2000 this sequence version replaced gi:7328772.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L7273
 Center clone name: 174_F10

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 134873 bases at least Q40
 Consensus quality: 142261 bases at least Q30
 Consensus quality: 145325 bases at least Q20
 Insert size: 155000; agarose-fp
 Insert size: 147529; sum-of-contigs
 Quality coverage: 3.7 in Q20 bases; agarose-fp
 Quality coverage: 3.9 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1572: contig of 1572 bp in length
 * 1573 1672: gap of 100 bp
 * 1673 2741: contig of 1069 bp in length
 * 2742 2841: gap of 100 bp
 * 2842 4870: contig of 2029 bp in length
 * 4871 4970: gap of 100 bp
 * 4971 6056: contig of 1086 bp in length
 * 6057 6156: gap of 100 bp
 * 6157 7419: contig of 1263 bp in length
 * 7420 7519: gap of 100 bp
 * 7520 9305: contig of 1786 bp in length
 * 9306 9405: gap of 100 bp
 * 9406 11199: contig of 1794 bp in length
 * 11200 11299: gap of 100 bp
 * 11300 13054: contig of 1755 bp in length
 * 13055 13154: gap of 100 bp
 * 13155 15917: contig of 2663 bp in length
 * 15918 15917: gap of 100 bp
 * 15919 18265: contig of 2348 bp in length

Db 70675 GAGATCTGTTAACCATCTGTGCTTACTTCACTGATGATCTAGCCAGTGTCTGTTCCCC 70734
QY 861 TTCCAAATTAAGGATGACGGAGTATAGGTAGAGACCCCTCCATCTCTGCGATTACAG 920
Db 70735 TTCCAAATTAAGGATGACGGAGTATAGGTAGAGACCCCTCCATCTCTGCGATTACAG 70734
QY 921 TTTTACGGTGTGATTCAGTAATCTCATCCAGGACAGATGATGCTCTGCTGTTAAAA 980
Db 70795 TTTTACGGTGTGATTCAGTAATCTCATCCAGGACAGATGATGCTCTGCTGTTAAAA 70854
QY 981 GCAAGAGAACTATATACCAAGATGAGGACCTCACTGCGAAATCAGCTCTCTAGGACAGA 1040
Db 70855 GCAAGAGAACTATATACCAAGATGAGGACCTCACTGCGAAATCAGCTCTCTAGGACAGA 70914
QY 1041 TCTAGGGAGTCCAAATTCCTTTTCCCAATGATGAGGGGATTTGATGAAAAAGAGCC 1100
Db 70915 TCTAGGGAGTCCAAATTCCTTTTCCCAATGATGAGGGGATTTGATGAAAAAGAGCC 70974
QY 1101 AGCAGAGGAGACACCACTGAAGAATCCCTCGGCTCTGTTTACCTTTGCTTCTCAAAACC 1160
Db 70975 AGCAGAGGAGACACCACTGAAGAATCCCTCGGCTCTGTTTACCTTTGCTTCTCAAAACC 71034
QY 1161 AGACATGCTTAATGGTCTGGAGGAATAATGACTGTGAACGGTGTTCAGATTGCTTGT 1220
Db 71035 AGACATGCTTAATGGTCTGGAGGAATAATGACTGTGAACGGTGTTCAGATTGCTTGT 71094
QY 1221 GCCTAATGAAGTTAGGGCTGATGAAGATGAAGTTATGAACATGAAGAACTCTTGGCAC 1280
Db 71095 GCCTAATGAAGTTAGGGCTGATGAAGATGAAGTTATGAACATGAAGAACTCTTGGCAC 71154
QY 1281 TACAGAATTCCTTAATATGACAGACATTTCTCTGAATCTCAGGACATGACTAATTTGAA 1340
Db 71155 TACAGAATTCCTTAATATGACAGACATTTCTCTGAATCTCAGGACATGACTAATTTGAA 71214
QY 1341 GTTGACTAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db 71215 GTTGACTAACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 71274
QY 1401 ACAGATTAGTCAGGCTGAGGACACTAATGTTGATGATGAGGACAGTGTGTTGATTTGGC 1460
Db 71275 ACAGATTAGTCAGGCTGAGGACACTAATGTTGATGATGAGGACAGTGTGTTGATTTGGC 71334
QY 1461 AGATCAGGCTTAGATTAAAGGAACCTTGATGATGATGATGATGATGATGATGATGATGAT 1520
Db 71335 AGATCAGGCTTAGATTAAAGGAACCTTGATGATGATGATGATGATGATGATGATGATGAT 71394
QY 1521 CACTGTTATAGACACACAGGAGGAGTGTATCTATCTAATGTTGATGATGATGATGATGAT 1580
Db 71395 CACTGTTATAGACACACAGGAGGAGTGTATCTAATGTTGATGATGATGATGATGATGAT 71454
QY 1581 GCAAGACCCAGGTTCTTTTGTTCGAAAGACTTTTACCTTCCAAAGAGATTTCAGTAAC 1640
Db 71455 GCAAGACCCAGGTTCTTTTGTTCGAAAGACTTTTACCTTCCAAAGAGATTTCAGTAAC 71514
QY 1641 AGAAGAAAGAAATAGGAGGAGGAGTGTATCTAATGTTGATGATGATGATGATGATGATGAT 1700
Db 71515 AGAAGAAAGAAATAGGAGGAGGAGTGTATCTAATGTTGATGATGATGATGATGATGATGAT 71574
QY 1701 AGGAAATGAGGCCACAGAGGAGGAGTGTATCTAATGTTGATGATGATGATGATGATGATGAT 1760
Db 71575 AGGAAATGAGGCCACAGAGGAGGAGTGTATCTAATGTTGATGATGATGATGATGATGATGAT 71634
QY 1761 GAAATATATTTACATATTTCTAGTCAAGTTCATCATGCTGTTGGGCAATCTTCCCC 1820
Db 71635 GAAATATATTTACATATTTCTAGTCAAGTTCATCATGCTGTTGGGCAATCTTCCCC 71694
QY 1821 CAAGGTAGTAGCAAGCTGCGCATCTATCATGTTGTTTCTTTTGGTGGTGAAGACCCCAAGCA 1880
Db 71695 CAAGGTAGTAGCAAGCTGCGCATCTATCATGTTGTTTCTTTTGGTGGTGAAGACCCCAAGCA 71754
QY 1881 ACCTTCTAATCTTAACTTCAATTCGAAAGCCATATGACACATTTCAAAATGACTTT 1940

Db 71755 ACCTTCTAATCTTAACTTCAAAATTCGAAAGCCATATCAGACCATTTACAAAATGACTT 71814
QY 1941 TCTTGCACACAGTGGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2000
Db 71815 TCTTGCACACAGTGGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 71874
QY 2001 AGGGGAAATCTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2060
Db 71875 AGGGGAAATCTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 71934
QY 2061 TACAAATGGGAAATCTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTTCCG 2120
Db 71935 TACAAATGGGAAATCTTAGAAAGTTATGAGGCTGAGATCTCCACTAGACCATGCTTCCG 71994
QY 2121 ATTAGTCTCAGATAGCCAGATAATGATCTCAGAGCTGCTCAGTTTGGAAATTTCTGCCAG 2180
Db 71995 ATTAGTCTCAGATAGCCAGATAATGATCTCAGAGCTGCTCAGTTTGGAAATTTCTGCCAG 72054
QY 2181 AAAGCCATTCACCACTCGGTGAGTGTGCTCAGTATGGTACCGGATTTCTCAGGCTCC 2240
Db 72055 AAAGCCATTCACCACTCGGTGAGTGTGCTCAGTATGGTACCGGATTTCTCAGGCTCC 72114
QY 2241 AAATTCATGAATGTGAAGCCAGGTTTACATTCACCAAAAGGAGGCATCCTGCGAGAGC 2300
Db 72115 AAATTCATGAATGTGAAGCCAGGTTTACATTCACCAAAAGGAGGCATCCTGCGAGAGC 72174
QY 2301 ATGTGGGAGGT 2312
Db 72175 ATGTGGGAGGT 72186

RESULT 7
AX809313
LOCUS AX809313
DEFINITION Sequence 10 from Patent WO03045990.
ACCESSION AX809313
VERSION AX809313.1 GI:38523628
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Legrain, P., Gauthier, J.M., Colland, F. and Jacq, X.
TITLE Protein-protein interactions involving transforming growth factor
beta signalling
JOURNAL Patent: WO 03045990-A 10 05-JUN-2003;
Hybridomics (PR)
FEATURES
Location/Qualifiers
1..1977
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 40.9%; Score 1977; DB 6; Length 1977;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2434 AGTGCTCAAGCCAGAGCCCTAACCTCAACAACTCTCTGTAATCTGTTCTACTATCCCT 2493
Db 1 AGTGCTCAAGCCAGAGCCCTAACCTCAACAACTCTCTGTAATCTGTTCTACTATCCCT 60
QY 2494 CCTTGGAGAGCTCAGGCTCAGGCTCTAGCTCTCCACTCCCACTGTGATGATA 2553
Db 61 CCTTGGAGAGCTCAGGCTCAGGCTCTAGCTCTCCACTCCCACTGTGATGATA 120
QY 2554 COTGTGGAGGTTTAAAGCACCTTGGAGCAAGTGTCTAGCCAGAGCAGAGCGGA 2613
Db 121 COTGTGGAGGTTTAAAGCACCTTGGAGCAAGTGTCTAGCCAGAGCAGAGCGGA 180
QY 2614 GTTGGTTGCTGATGGATCTTCCCAATGGAGAGTGTCTGATGAGCAAAATTAACA 2673

Db	181	GTCTGGTTGCTGATGGATCTTCCCAATGGAGAAAGTTGCTGATGACGCCAAATTAACA	240
Qy	2674	ATGAATGAATCTTCTCTGAGAAACCTGGCTGTGTACACACCCAGTCAAGCCAGTA	2733
Db	241	ATGAATGAATCTTCTCTGAGAAACCTGGCTGTGTACACACCCAGTCAAGCCAGTA	300
Qy	2734	ACTACCACTCTCTACACAGAGACGATATTTGTTCTATTTCTGGGAGTATAAATCAG	2793
Db	301	ACTACCACTCTCTACACAGAGACGATATTTGTTCTATTTCTGGGAGTATAAATCAG	360
Qy	2794	GTTGGAAAGTCTGTGGAAGTGAATGAATTTATTTCTGAAGATGGCTTCTCCCAT	2853
Db	361	GTTGGAAAGTCTGTGGAAGTGAATGAATTTATTTCTGAAGATGGCTTCTCCCAT	420
Qy	2854	CTCATCTCCATCGGTGTAAGAGAGACTATGCTGTGGAAGAGAAACCATCACAGATTCA	2913
Db	421	CTCATCTCCATCGGTGTAAGAGAGACTATGCTGTGGAAGAGAAACCATCACAGATTCA	480
Qy	2914	GTAATGACGAGTTGGAGGATGTTGGCCCTGACCCATTTGTTTAAATGCAAT	2973
Db	481	GTAATGACGAGTTGGAGGATGTTGGCCCTGACCCATTTGTTTAAATGCAAT	540
Qy	2974	TTGTTGTCATGTTTAAATTTGTAATTTATGTAACAGGAGTCTGGTGTTCACAACC	3033
Db	541	TTGTTGTCATGTTTAAATTTGTAATTTATGTAACAGGAGTCTGGTGTTCACAACC	600
Qy	3034	AAGGGAATGATGATGAGTGGGTGAGTCTGAGATAGTCTTCTTACAGTGTTCACCGAT	3093
Db	601	AAGGGAATGATGATGAGTGGGTGAGTCTGAGATAGTCTTCTTACAGTGTTCACCGAT	660
Qy	3094	GAAGAAGTGTTCGCAAGGATATCTTTAATCACTTTGTGAGCTTTATCGGATGCTCTG	3153
Db	661	GAAGAAGTGTTCGCAAGGATATCTTTAATCACTTTGTGAGCTTTATCGGATGCTCTG	720
Qy	3154	GCAGGAATGTTGAGCACTCGGACATTCCTTCTTCACTCAAGATTTTCTTGGCAGT	3213
Db	721	GCAGGAATGTTGAGCACTCGGACATTCCTTCTTCACTCAAGATTTTCTTGGCAGT	780
Qy	3214	AAAGAACATGTTGATTTCTTATATGTGACATCTACCTACCACTCACTGCAAGACTAGTA	3273
Db	781	AAAGAACATGTTGATTTCTTATATGTGACATCTACCTACCACTCACTGCAAGACTAGTA	840
Qy	3274	CTCCCAACCCACCTTACTTGTGTTGGATCTTATCCAGAAATCGGAACCTCCTTGGGCT	3333
Db	841	CTCCCAACCCACCTTACTTGTGTTGGATCTTATCCAGAAATCGGAACCTCCTTGGGCT	900
Qy	3334	AAAGTATTTCTATCCGCTGATGTTGAGACTTGGAGCTGAATATCGACTTTATCCATGC	3393
Db	901	AAAGTATTTCTATCCGCTGATGTTGAGACTTGGAGCTGAATATCGACTTTATCCATGC	960
Qy	3394	CCACTATTCAGTTCAGATTTGGAGCCATTTGTTGGAGAGCGGGGCATACCATCATG	3453
Db	961	CCACTATTCAGTTCAGATTTGGAGCCATTTGTTGGAGAGCGGGGCATACCATCATG	1020
Qy	3454	ATCTCTCTGAGACTTCAGAAATACAGATATACCTTCCAGTAGTTCAGAGTTTGGTG	3513
Db	1021	ATCTCTCTGAGACTTCAGAAATACAGATATACCTTCCAGTAGTTCAGAGTTTGGTG	1080
Qy	3514	GTTGATGGAAGTTCCGAAACTAGCATCAAAATTCACCAACAGATACAATGAGATG	3573
Db	1081	GTTGATGGAAGTTCCGAAACTAGCATCAAAATTCACCAACAGATACAATGAGATG	1140
Qy	3574	ATGAAGCCATGAACAAGTCCATGAGCATGTCCTGGCAGGAGTGCCTGCTTCAATGAA	3633
Db	1141	ATGAAGCCATGAACAAGTCCATGAGCATGTCCTGGCAGGAGTGCCTGCTTCAATGAA	1200
Qy	3634	AAGGAGACTCTCATCTTGTGTGTACAGATATGATGGAACCTATCAGACCCAGGCT	3693
Db	1201	AAGGAGACTCTCATCTTGTGTGTACAGATATGATGGAACCTATCAGACCCAGGCT	1260
Qy	3694	ATCAGTATTCAAATCAGCCCAAGAAAGTACTGTTGCTGCTTCTTGTGTTGCTGCTGC	3753
Db	1261	ATCAGTATTCAAATCAGCCCAAGAAAGTACTGTTGCTGCTTCTTGTGTTGCTGCTGC	1320

Qy	3754	GCTCTGAAATCCTCTTCTGGATACCTTCCCAAGTCCAGTATTGTTGGAAGATGGTGTATG	3813
Db	1321	GCTCTGAAATCCTCTTCTGGATACCTTCCCAAGTCCAGTATTGTTGGAAGATGGTGTATG	1380
Qy	3814	GTCAGATTTACTGCAGAGAACATGGATTCTTGTGAGGAGGACACTGCGAGAGATGAAGAC	3873
Db	1381	GTCAGATTTACTGCAGAGAACATGGATTCTTGTGAGGAGGACACTGCGAGAGATGAAGAC	1440
Qy	3874	TTCCACCATCACCTGTGGGAAGCGGACCGGAGGAACCCCGAGGAGCACATCCACATCCAG	3933
Db	1441	TTCCACCATCACCTGTGGGAAGCGGACCGGAGGAACCCCGAGGAGCACATCCACATCCAG	1500
Qy	3934	TGGTGGATGATGACAAGAACCGTTAGCAAGGGTGTCTAGTCTTATAGATGGGAGTCC	3993
Db	1501	TGGTGGATGATGACAAGAACCGTTAGCAAGGGTGTCTAGTCTTATAGATGGGAGTCC	1560
Qy	3994	ATGGAGACTATACAAATGTGAAGATATTCATGGATCAGATATATAAGCAAAATGGAATA	4053
Db	1561	ATGGAGACTATACAAATGTGAAGATATTCATGGATCAGATATATAAGCAAAATGGAATA	1620
Qy	4054	GTAATCAGATGACAGAGAGTGTTCCTAGAAAAAGATGACCAAACTGGGACTACGTGTGACA	4113
Db	1621	GTAATCAGATGACAGAGAGTGTTCCTAGAAAAAGATGACCAAACTGGGACTACGTGTGACA	1680
Qy	4114	GATCTGAGATCAGATAGATTGACTGAGCATGTTGCCAAAGCTTTTTCCTTGTCTCTC	4173
Db	1681	GATCTGAGATCAGATAGATTGACTGAGCATGTTGCCAAAGCTTTTTCCTTGTCTCTC	1740
Qy	4174	TGTCTCTCACTCAAACTTCTGAAGGAAGATGAAATGACCAAACTGGGACTACGTGTGACA	4233
Db	1741	TGTCTCTCACTCAAACTTCTGAAGGAAGATGAAATGACCAAACTGGGACTACGTGTGACA	1800
Qy	4234	CTTGACTCAGATCAGTGTGGCTATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAG	4293
Db	1801	CTTGACTCAGATCAGTGTGGCTATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAG	1860
Qy	4294	TACATCAATGATCTGATAGCGCTTGTGCGCGTGTATCCATGGAGGGGCTTCCAGCTT	4353
Db	1861	TACATCAATGATCTGATAGCGCTTGTGCGCGTGTATCCATGGAGGGGCTTCCAGCTT	1920
Qy	4354	AGTGGGGGGGGCTTCTCATGGAATCATCTTTTATATTTCTGGAATAACATCGATATA	4410
Db	1921	AGTGGGGGGGGCTTCTCATGGAATCATCTTTTATATTTCTGGAATAACATCGATATA	1977

RESULT 8
AL591398
LOCUS
DEFINITION Human DNA sequence from clone RP11-81C12 on chromosome X, complete sequence.
ACCESSION AL591398
VERSION AL591398.2 GI:14330165
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 89973)
Heath, P.
Direct Submission
Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14133279.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-81C12 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

IMPORTANT: This sequence is not the entire insert of clone Rpl1-91C. It may be shorter because the sequence overlapping sections only end of clone Rpl1-389P13 is at 83972. The true left end of clone Rpl1-389P13 is at 83972. The true right end of clone Rpl1-137H15 is at 100 in this sequence. The true right end of clone Rpl1-7706 is at 44291 in this sequence. Location/Qualifiers
 1...89973

FEATURES

```

1: 89573
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-81C12"
/clone_lib="RPCT-11.1"
1: 791
/notes="LIM2 repeat: matches 5524..6308 of consensus"
914..1079
/notes="LIMD repeat: matches 994..1160 of consensus"
1130..1329
/notes="LIMEC repeat: matches 2218..2411 of consensus"
1330..1377
/notes="MLT1a1 repeat: matches 318..365 of consensus"
1378..1680
/notes="AluJo repeat: matches 1..308 of consensus"
1681..1848
/notes="MLT1a1 repeat: matches 155..318 of consensus"
1849..2152
/notes="AluSx repeat: matches 1..299 of consensus"
2153..2305
/notes="MLT1a1 repeat: matches 1..155 of consensus"
2306..2602
/notes="LIMEC repeat: matches 1901..2218 of consensus"
2642..3475
/notes="LIMEC repeat: matches 1245..2090 of consensus"
3476..3860
/notes="THE1C repeat: matches 3..371 of consensus"
3861..3929
/notes="LIMEC repeat: matches 2090..2159 of consensus"
3930..4240
/notes="AluY repeat: matches 1..306 of consensus"
4241..4885
/notes="LIMEC repeat: matches 2159..2430 of consensus"
4886..5198
/notes="AluDb repeat: matches 1..309 of consensus"
5199..5491
/notes="LIMEC repeat: matches 2430..2726 of consensus"
5491..5917
/notes="LIM4 repeat: matches 3146..3578 of consensus"
5918..6235
/notes="AluDb repeat: matches 1..312 of consensus"
6236..6976
/notes="LIM4 repeat: matches 3578..4297 of consensus"
7047..7160
/notes="Charlie1 repeat: matches 2022..2140 of consensus"
7161..7262

```

repeat_region	/note="LMC3 repeat: matches 7440. .7542 of consensus" 7263. .7581
repeat_region	/note="Alusx repeat: matches 1. .302 of consensus" 7592. .7776
repeat_region	/note="LMC3 repeat: matches 7542. .7734 of consensus" 7777. .7791
repeat_region	/note="Charliel repeat: matches 2140. .2153 of consensus" 8110. .8960
repeat_region	/note="Charliel repeat: matches 2610. .2761 of consensus" 9033. .9458
repeat_region	/note="Charliel repeat: matches 2190. .2627 of consensus" 9506. .9826
repeat_region	/note="Alusg repeat: matches 1. .308 of consensus" 9938. .10017
repeat_region	/note="Alusx repeat: matches 1. .80 of consensus" 10019. .10166
repeat_region	/note="LMC3 repeat: matches 6611. .6747 of consensus" 10205. .10502
repeat_region	/note="Alusx repeat: matches 1. .311 of consensus" 10538. .10681
repeat_region	/note="LMC3 repeat: matches 6468. .6609 of consensus" 11563. .11712
repeat_region	/note="FAM repeat: matches 13. .165 of consensus" 11906. .12226
repeat_region	/note="Alusx repeat: matches 1. .312 of consensus" 12345. .12852
repeat_region	/note="Alusg repeat: matches 1. .310 of consensus" 12992. .13267
repeat_region	/note="Alubg repeat: matches 1. .285 of consensus" 13268. .13327
repeat_region	/note="30 copies 2 mer ta 98% conserved" 14758. .14984
repeat_region	/note="MIR repeat: matches 16. .257 of consensus" 15103. .15315
repeat_region	/note="L2 repeat: matches 1551. .1774 of consensus" 15324. .15625
repeat_region	/note="L2 repeat: matches 2442. .2750 of consensus" 16017. .16079
repeat_region	/note="MIR repeat: matches 78. .139 of consensus" 16085. .16215
repeat_region	/note="MER63 repeat: matches 704. .1687 of consensus" 16594. .16617
repeat_region	/note="12 copies 2 mer tg 95% conserved" 16923. .16980
repeat_region	/note="29 copies 2 mer aa 74% conserved" 17294. .17497
misc_feature	/note="MIR repeat: matches 12. .254 of consensus" 17730. .18153
misc_feature	/note="match: GSS: Em:AQ264531" 17743. .17945
repeat_region	/note="match: GSS: Em:AQ263645" 18255. .18278
repeat_region	/note="12 copies 2 mer gt 100% conserved" 18578. .18872
repeat_region	/note="Alubd repeat: matches 1. .294 of consensus" 19101. .19464
repeat_region	/note="MER39 repeat: matches 13. .388 of consensus" 19508. .19598
repeat_region	/note="L2 repeat: matches 2662. .2750 of consensus" 20288. .20583
repeat_region	/note="Alubd repeat: matches 1. .293 of consensus" 21186. .21309
repeat_region	/note="62 copies 2 mer tt 59% conserved" 21665. .21708
repeat_region	/note="22 copies 2 mer ac 77% conserved" 23271. .23419
repeat_region	/note="MIR repeat: matches 37. .173 of consensus" 23420. .23720
repeat_region	/note="Alusx repeat: matches 3. .310 of consensus" 23721. .23800
repeat_region	/note="MIR repeat: matches 173. .255 of consensus" 23906. .23960
repeat_region	/note="M12B repeat: matches 394. .448 of consensus"

repeat_region	23961..24050	/notes="45 copies 2 mer ta 88% conserved"	
repeat_region	24182..24576	/notes="MLT2B repeat: matches 1..404 of consensus"	
repeat_region	25118..25401	/note="LIME1 repeat: matches 5747..6063 of consensus"	
misc_feature	28019..28534	/notes="match: GSS: Em:AZ519577"	
repeat_region	28603..28695	/notes="MIR repeat: matches 52..146 of consensus"	
repeat_region	29276..29754	/notes="L2 repeat: matches 1891..2393 of consensus"	
repeat_region	29933..30218	/notes="AluX repeat: matches 1..286 of consensus"	
repeat_region	31138..31447	/notes="AluS repeat: matches 1..309 of consensus"	
repeat_region	32120..32398	/notes="MER7A repeat: matches 62..346 of consensus"	
repeat_region	32401..32510	/note="55 copies 2 mer ta 72% conserved"	
repeat_region	32533..32560	/notes="14 copies 2 mer ac 100% conserved"	
repeat_region	32561..32635	/notes="78K repeat: matches 1..73 of consensus"	
repeat_region	33629..33658	/note="15 copies 2 mer ct 90% conserved"	
repeat_region	34128..34286	/note="MIR repeat: matches 95..262 of consensus"	
repeat_region	34349..34684		
Query Match			32.1%; Score 1552.2; DB 9; Length 89973;
Best Local Similarity			94.1%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 83; Indels 21; Gaps 3;			
Qy	2640	CAATGGAGAGTTCTGATGAGCCAAATTAACAAATGAATGGAACCTCTCCAGGAAC	2699
Db	51977	CAAAAGAGAGTTCTGATGAGCCAAATTAAGCAATGAATGGAACCTCTCCAGGAAC	52036
Qy	2700	CCTGGCTGTGCACACACCCAGTCAAGCCAGTAACCTACCAGTCTCTACAGCAGAGAC	2759
Db	52037	CCTGGCTGTGTCAACACACCCAGTCAAGCTAGTAACCTACCAGTCTCTACAGCAGAGAT	52096
Qy	2760	GGATATTGTCTATCTCTGGAGATTAATCTAGGTTGGAAGTCTCTTTGGAAGTGAAT	2819
Db	52097	GGATATTCTCTATCTCTGGAGATTAATCTAGGTTGGAAGTCTCTTTGGAAGTGAAT	52156
Qy	2820	GAATCTTATCTGAAGATGCTCTCTCCATCTCATCTCCAGTGTGTAAAGGAGA	2879
Db	52157	GAGCTTATCTGAAGATGCTCTCTCCATCTCATCTCCAGTGTGTAAAGGAGA	52216
Qy	2880	CTATGCTGTGAAGAGAAACCATCACAG- - -ATTTCAGTAATGACAGTTGGAGATG	2935
Db	52217	CTATGCTGTGAAGATAAACCATCACAGATTTATTTTCAGAAATGCAGAGTTGGAAGACG	52276
Qy	2936	GTGGCCCTGACCCATGTTATTTTAAATGCAAAATTTGTTCTCAATGGTTAAATTTG	2995
Db	52277	GTGGCCCTGACCCATGTTATTTTAAATGCAAAATTTGTTCTCAATGGTTAAATTTG	52336
Qy	2996	TAAATATGTGAACAGGAAGTGTGTTTTCACAAACCAAGGAATGCATGCGGTG	3055
Db	52337	TAAATATGTGAACAGGAAGTGTGTTTTCACAAACCAAGGAATGCATGCGGTG	52396
Qy	3056	AGCTCGAGATGATCTATCTTACAGTGTTTACCGGATGAAAGTGTGTTGCAAGGATA	3115
Db	52397	AGCTCGAGATGATCTATCTTACAGTGTTTACCGGATGAAAGTGTGTTGCAAGGATA	52456
Qy	3116	TCTTTAATCACTTGTGAGTGTATCGGATGCTCTGGCAGGAATGTGTTGAGCAACT	3175
Db	52457	TCTTTAATCACTTGTGAGTGTATCGGATGCTCTGGCAGGAATGTGTTGAGCAACT	52515
Qy	3176	TGGGACATTCCTTCTAGTCAAAAGTTTCTTGGCAGTAAAGACATGGTGATTTCTAT	3235
Db	52516	TGGGACATTCCTTCTAGTCAAAAGTTTCTTGGCAGTAAAGACATGGTGATTTCTAT	52575

Qy	3236	ATGTGACATCTACCTACAGTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTGT	3295
Db	52576	ATGTGACATCTACCTACAGTCACTGCAAGACCTAGTACTCCCAACCCACCTTACTTGT	52635
Qy	3296	TTGGGATTCCTTATCCAGAAATGGGAAACTCCTTGGGTAAAGTATTTCTTATCCGTCTGA	3355
Db	52636	TTGGGATTCCTTATCCAGAAATGGGAAACTCCTTGGGTAAAGTATTTCTTATCCGTCTGA	52695
Qy	3356	TGTTGAGACTTGGAGCTGAATATCGACTTTATCCATGCCCACTATTAGTGTGAGATTTC	3415
Db	52696	TGTTGAGACTTGGAGCTGAATATCGACTTTATCCATGCCCACTATTAGTGTGAGATTTC	52755
Qy	3416	GGAAAGCCATGTTTGGAGAGACGGGCATACCATCATGAATCTTCTTTCAGACTTCAGAA	3475
Db	52756	GGAAAGCCATTATTTGGAGAGACGGGCATACCATCATGAATCTTCTTTCAGACTTCAGAA	52815
Qy	3476	ATTACAGATATACCTTGGCAGTAGTTCAAGTTTGGTGGTTGATATGAAAGTTCGGAATA	3535
Db	52816	ATTACAGATATACCTTGGCAGTAGTTCAAGTTTGGTGGTTGATATGAAAGTTCGGAATA	52875
Qy	3536	CTAGCATCAAAATTCAGCAACACAGATACATGAGATGATGAAAGCCATGAACAGTCCA	3595
Db	52876	CTAGCATCAAAATTCAGCAACACAGATACATGAGATGATGAAAGCCATGAACAGTCCA	52935
Qy	3596	ATGAGCATGCTCTGCGCAGGAGTGCCTCTCAATGAAGAAAGCAGACTCTCATCTTGTGT	3655
Db	52936	ATGAGCATGCTCTGCGCAGGAGTGCCTCTCAATGAAGAAAGCAGACTCTCATCTTGTGT	52995
Qy	3656	GTGTACAGATGATGATGGAACATACAGACCCAGCTATCAGTATTCACATCAGCCCA	3715
Db	52996	GT- - - - -GAGCTATCAGACCCAGCTATCAGTATTCACATCAGCCCA	53039
Qy	3716	GAAGAGTGACTGCTGCCAGTTTCTTTGTGTTGAGTGGGCTCTGAAATCTCTTCTGTGT	3775
Db	53040	GAAGAGTGACTGCTGCCAGTTTCTTTGTGTTGAGTGGGCTCTGAAATCTCTTCTGTGT	53099
Qy	3776	ACCTTGCAGTCCAGTATTTGGAAGATGTTTATGTTTCCAGTATTCAGAGAGAAC	3835
Db	53100	ACCTTGCAGTCCAGTATTTGGAAGATGTTTATGTTTCCAGTATTCAGAGAGAAC	53159
Qy	3836	TGATTTCTTTGAGGAGGACCTGCGAGAGATGAAGAGCTTCCACATCCTCTGGGAAGG	3895
Db	53160	TGATTTCTTTGAGGAGGACCTGCGAGAGATGAAGAGCTTCCACATCCTCTGGGAAGG	53219
Qy	3896	CGAGCCGAGGAAACCCAGGAGCAGTCCCATCCAGTGGGTGATGATGACAGAGACG	3955
Db	53220	CGAGCCGAGGATCCCGAGGAGCAGTCCCATCCAGTGGGTGATGATGACAGAGACG	53279
Qy	3956	TTAGCAAGGGTGTCTGTAAGTCTTATAGATGGGAAGTCCATGGAGACTATAACAAATGTGA	4015
Db	53280	TTAGCAAGGGCGTCTGTAAGTCTTATAGATGGGAAGTCCATGGAGACTATAACAAATGTGA	53339
Qy	4016	AGATATTTCCATGGATCAGAAATAAAGCAAAATGGAAAGTAAATCAGATGGAGAGTGT	4075
Db	53340	AGATATTTCCATGGATCAGAAATAAAGCAAAATGGAAAGTAAATCAGATGGAGAGTGT	53399
Qy	4076	TTTTCTAGAAACAGATGACAGCAGCAATTCCTCAGTGCATCTGCGAGATCAGTAGAT	4135
Db	53400	TTTTCTAGAAACAGATGACAGCAGCAATTCCTCAGTGCATCTGCGAGATCAGTAGAT	53459
Qy	4136	TGACTCAGCATGTTGCAAAAGTCTTTTGGCTTGTCTCTGTCTCTCACTCTGAAATCTCTGA	4195
Db	53460	TGACTCAGCATGTTTACCAAGGCTTTTGTCTCTCTGTCTCTCACTCTGAAATCTCTGA	53519
Qy	4196	AGAAAGATGGAATGACAAACTCGGAGTACGTGTGACATTCAGTACAGATCAGTTGGCT	4255
Db	53520	AGAAAGATGGAATGACAAACTCGGAGTACGTGTGACATTCAGTACAGATCAGTTGGCT	53579
Qy	4256	ATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGAATGATCTGGATAGG	4315
Db	53580	ATCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGAATGATCTGGATAGG	53639

```

QY 4316 CCTTGGTCCGGTATCCATGAGGGGCTCCACCTTAGTGTAGGAGGCCCGCTGTGTCTATGG 4375
DB 53640 CCTTGGTCCGGTATCCATGAGGGGCTGTGTAGTGTAGGAGGCCCGCTGTGTCTATGG 53699
QY 4376 AACTCATCTTTTATATT 4392
DB 53700 AACTCATCTTTTATTTT 53716

RESULT 9
AF104305 4823.bp mRNA linear VRT 03-JAN-1999
LOCUS Xenopus laevis Smad anchor for receptor activation (SARA) mRNA,
DEFINITION complete cds.
ACCESSION AF104305
VERSION AF104305.1 GI:4092768
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 4823)
AUTHORS Tsukazaki,T., Chiang,T.A., Davison,A.F., Attisano,L. and Wrana,J.L.
TITLE SARA, a FIVE domain protein that recruits Smad2 to the TGFbeta
receptor
JOURNAL Cell 95 (6), 779-791 (1998)
MEDLINE 99081294
PUBMED 986596
REFERENCE 2 (bases 1 to 4823)
AUTHORS Wrana,J.L.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Program in Developmental Biology, Hospital
for Sick Children, 555 University Avenue, Toronto, Ontario M6G 1X8,
Canada

FEATURES
source
1..4823
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
1..4823
/gene="SARA"
316..4023
/gene="SARA"
/codon_start=1
/product="Smad anchor for receptor activation"
/protein_id="AAC99463.1"
/db_xref="GI:4092769"
/translation="MPKRVGDTIMAEDSLFPNNTGSPSEIVCNSTVESQSLQALDQDS
VNIHNEKSVLLADGFPSCSPKSLINFDCLITMONELPHSQMSVDNDKETVTSIVLP
TIIQDTSNDPAINKPGTKPEHRALKETTSVILPEIKPSTCAALSFNNNKVPY
QLNNTDLSVSVVACSEQQQKHTSLHEELKEFEGVSATPFAATAETVLNEALR
SAEPFDIVKRNFSQVINGDLTKSLGSLSENEKFCASKEFEGVDANVLENAVA
YKRAIDLPENGTNAPMSLVNGCDSYGMKNPAVAQNPKNLPKSDSVTEKETEESSK
EYTVGVYEQREDVTERTGILLNAKADQKNLHSLCNGVPSMSHQTSPKPKIVQS
LSVYGGARTQPHLKPHLPSEMLQSDLPNAGGSSKYKNDMLKNSQDNL
SELRDVSARVPVTDANGPFGYRGPGLSLAVSPDSPDNDLLAGQFQPSKDF
TLGEVAPVWPDSPQAPNCKEARFTTKRHRACRACGVCAACC SLXKLIQYMDK
EARCVI CHSVLNAQAWENLMSASVQSPNPNPABEYCTIPPMQQAQSGALSSPP
TMVPVGLVHPGTEGSKQORRVWFDGLPNEGTDSDNANVTVAAGTLTVSHTN
NNTSSEBNTSGFGSITQVGSAMNLI PEDGLPILISTGVKGDYAVEERPSQSMVQ
QLREGGPDPLAVFVLANLLAMKLVNVRKCHCFTTKGMHVGQREIVILQCLPDE
KLCPRDFSHFVLQIAAGNVVGNLHGSFLSQFLGSKDGHGFLYVAPTSIQDGL
VLPAEPLFGILIQKWTETPAKVPFIRLMRLGAELVPCPLFSVYRKLFGFTH
TIIINLADFRNYQTLFVQGLVVDMEVRKTSIKIPSNRYNEMKMNKSNHVLAI
ACFNQADSHLVQNDQNGYQTAIS IHKQPKVTGAPFPVPSGALKSSSGVLAKS
IVEDGVNVIQTASMDALROSLREMKDFTTCGKADAEBSQEHVYVQWVEDDKNFKG
VFSPIDKSNVESVSKI PRGSEYKASGKTIIRWIEVFELDNEEQSGLSDPADHFNLT
ENKAFACIALCPHLKILKEDGMTRGLGLVSLDSQVGVQAGSNGQLLPARTNDLDG
ALVPVTHGGTCQLSEGPVSMELFYILENIS"

31.9%; Score 1545.8; DB 5; Length 4823;

ORIGIN
Query Match

```

```

Best Local Similarity 71.8%; Pred. No. 0;
Matches 2136; Conservative 0; Mismatches 792; Indels 45; Gaps 7;

QY 1443 ACAGTGTGTGTGATTGGCAGATGTCAGGTCTAGATTATAAAGGAACCTGCATTAGTGAAG 1502
DB 1089 AGAGTTTGAAGGAGGGGTAGTCTAATGCTCTGTGTGGAATGATGTGTAGCTATTAA 1148
QY 1503 TGAAGAAATGTATTTCTCCACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGG 1562
DB 1149 AGAAGCAATAGATTTCCTTGAAGAAATGAACTAATGACCAATGTCTCTGTACAATGG 1208
QY 1563 TTGTGATTCCTATGGAAATCAAGACCCAGCTGTTCTTTTGTTCCTCAAGACTTTACCTC 1622
DB 1209 GTGTGATTCCTATGGAAATGAAACCCAGCCGTAGCTCAAAACCCCAAGAAATTTACCTTC 1268
QY 1623 CAAAGAAGATTGAGTAACAGAGAAAGAAATAGAGGAAAGCAAGTCAGAAATGCTACTC 1682
DB 1269 AAAGAAGATTCTGTGACAGAGAAAGAAATGAAGAAAGCAAGTCAGAAATCTATATAC 1328
QY 1683 AATATTATG---AACAGAGAGAAATGAGGCCACAGAGGAGTGTGACTACTTTTAAA 1739
DB 1329 TGGTGTATTATGAACAACAAGAGAGAGATGATGTTACAGAGAGAGTGTGACTTCTGTAAA 1388
QY 1740 CAGCACTGTGTACCTTAATGAAGAAATTTATTTACATATTTCTGTAGTCAAGTTCCTATC 1799
DB 1389 TGCTAAGGCTGACCAATGAAGAAACA---ATTGATAGTCTTGTATCAGTTCCTATC 1445
QY 1800 AGTGCTTGGGCAATCTTCCCAAGGTAGTAGCAAGCTGCGCATCTATCAGTGTTCCTTT 1859
DB 1446 CATGATGGGCAACATCACCAAAAGGCAAGATTGTGCAATCTCTCAGTGTTCCTATA 1505
QY 1860 TGTGTGTGCAAGACCCCAAGACCTTCTAATCTTAACTTAACTTCAATTCCTCAAGCCATATC 1919
DB 1506 CGGTGAGCAGCCCATAGACGCCCAACTCTCTCAACTCTCAATTCCTCAAGCCATATC 1565
QY 1920 AGACCAATTTACAAA---ATGACTTTCTGCAAAAGTGTGAAATATATCTAAAAATAAAA 1976
DB 1566 TGAATATTGTGAGAGCGATCTCATCTCTCAATGCTGGTGCAGCTCTAAATACAAAA 1625
QY 1977 TGATATTCTTGGGAAGCAAAATAGGGGAAACTCAGCAACCATGATGTATGAGTCCATC 2036
DB 1626 TGACATGTTAAACAAATCAAAATCAGGGGGATAACCTGATTTTCAAGAACTCCTGCGTGAGGA 1685
QY 2037 TTTGGGAAACATCTCTAATGTGATACAAATGGGGAACATTTAGAAAGTTATGAGGCTGA 2096
DB 1686 TTCTGAGTGGCAGCCCTGT---TACTGATGCTAATGGTGTATTCCTCGAGAAATACAG 1742
QY 2097 GATCTCCACTAGACCATGCTTGCATTAGTCTCCAGATAGCCAGATAATGATCTCAGAGC 2156
DB 1743 GGGACCTGGCAGCTTGTGCTTGTGAGTCTCTCCAGACAGCCAGACACGATCTGCTTGC 1802
QY 2157 TGGTCAGTTTGGAAATTTCTGCGCAAGAACCAATTCACACCGCTGGGTGAGTGGCTCCAGT 2216
DB 1803 CGGGCAGTTTGGGTACCCATCTCTAAGCCATTTACTACTTAGGGNAAGTGGCTCCAGT 1862
QY 2217 ATGGTACCGGATTTCTAGGCTCCAAATGATGAATGTGAACCGAGTTTATCATTCAC 2276
DB 1863 CTGGGTGCGAGATTCCCAAGCACCACAACTGATGAAGTGCAGGCGCAGATTTACATTTAC 1922
QY 2277 CAAAAGGAGGATCATCTCAGAGCATGTGGGAAGTGTTCGTGCTTCTCTCTCTAGCTCT 2336
DB 1923 CAAAAGGAGGATCATCTCAGGAGCTTGTGGAAAGTGTTCGTGCTGCTGCTGCTGCT 1982
QY 2337 GAAATGTAACTGTTTATACATGGACAGAAAGAGTAGAGTGTGTGTATCTCCCAATC 2396
DB 1983 AAAATGCAACTACAGTACATGGATATAAAGAGAGCTGCTGTGTGTGTTATTCTCATTC 2042
QY 2397 AGTGTATGATCTCAGCTCAGCTGGGAAACATGATGATGCTCAAGCCAGAGCCCTAA 2456
DB 2043 TGTGTTATGATGCTCAGCATGGGAGACATGTTTAGTGTGATGCTGCTGCTGCTGCT 2102
QY 2457 CCCTAACCAATCTGCTGAATACTTCTTACTATATCCCTCCCTTGCAGCAAGCTCAGGCTC 2516

```


JOURNAL

Submitted (06-SEP-2002) Wellcome-Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

COMMENT
hummerY@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 12, 2002 this sequence version replaced ci:21212224.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Swi; SWISSPROT; Tr; TREMBL; Wp; WORMPEP. Information on the WORMPEP database can be found at

decrease can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-446P20 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

FEATURES

source

```

1. .85802
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone=RP23-446F20"
/clone_lib="RPCI-23"

```

ORIGIN

Query Match	29.4%; Score 1421.6; DB 10; Length 95802;
Best Local Similarity	82.3%; Pred. No. 0;
Matches 1739; Conservative	0; Mismatches 289; Indels 84; Gaps 6;
QY	201 AGATGAACAGTTCTCTCTACCTTTTATGGATCAAAAGTGGAAATAGATCTTAGATCCCCC 260
Db	AGATGAACAGTTCTCTCTACCTTTTATGGATCAAAAGTGGAAATAGATCTTAGATCCCTC 60599
QY	261 TTCTCACCGGCTGTCATTTAACCCCTACTTTGGCCCACTGTGAATGAATCTGCAGTTCTTAA 320
Db	TTCTCACCCATTGTCCTTTTAAACCCCTGCATTAGCCAGTGTGAATGAACCTACAGTTTC-- 60542
QY	321 TGAGTCACAAACACAACTGAAGTCTTCTCCCTGCTCATTCAGCTCCCTGACACACAGA 380
Db	TGAGACAGGACCACACGCTGAAGTCTTCTCTGGCTGGCTCAGCTCCCTTTACAAAGGA 60482
QY	381 GGAAGAGGATCACTGTGCTTAATGGAACAGACTGTGAATCTAAATCCAGAGATTGCCACAAT 440
Db	GGACAAAGATCCCTGTGCTTAACGGACAGGACTGTAGTCTAAATCCAGAGACTGACACCAT 60422
QY	441 GTGGATTGATGAAATGCTGTTGTCAGAGACCCAGTTAAATTAAAGAAACTATAGTTGGA 500
Db	GTGGATTGATGAAATGCTGTCACAGAGACCCAACTAAATTAAAGAAACTATAATCAGA 60362
QY	501 TGATCAATGCAGTGTGTTTGAAGTCGGAGAGAGAAATGTGGAACCTGGCTTGTCTGCC 560
Db	TGATCAATTTAGTCCGCTTGAAGTCGGAGAGAGAAATGTGAAGCTTCATTTGTCTGCC 60302
QY	561 AGATGAGAGAAATGTTCTTGTGTGACCGCTCATGCAATAACTGTGATAAAGAGCAATTACA 620
Db	AGA CAGAAGAAATGTTCTGGTTGTAGCTGTATGCACTAACTGCCATTAAGAGCAATTGCA 60242

Tue May 18 15:16:58 2004

```
Qy 1701 AGGAAATGAGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGGTGAACCTAATGAA 1760
Db 59239 AGGAAACAGAAATGCAAGAGGAGTGGACTGCTTTTAAACAGCACTGGGAATGAAATGAA 59180
Qy 1761 GAAAAATATTACATATTTCTGTAGTCAAGTTCCATCATCATGCTTGGGCAATCTCCCC 1820
Db 59179 GAAAAATATTATTACATATTTCTGTAGTCAAAATCCCATCATCATGCTTGGGCAATCTCCCC 59120
Qy 1821 CAAGTGTAGTACAGAGCTGGCATCTATCATGTGTTTCTTTTGGTGGTGGTCAAGCCCAAGCA 1880
Db 59119 CAAG--ATAGCAACCTGCAATCTATTAGTGTCTTTTGGTGGTGGTCAAGCCCAAGCA 59063
Qy 1881 ACCTTCTAATCTTAACTTCAAACTTCAAACTTCAAACTTCAAACTTCAAACTTCAAACTT 1940
Db 59062 ACCTTCTAATCTTAACTTCAAACTTCAAACTTCAAACTTCAAACTTCAAACTTCAAACTT 59003
Qy 1941 TCCTGCAACAGTGGAAATATATCTAAATAATAATAATAATAATAATAATAATAATAATAATA 2000
Db 59002 TCCTGCAACATAGGAATATATAGTAATAATAATAATAATAATAATAATAATAATAATAATA 58943
Qy 2001 AGGGGAAACTTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 2060
Db 58942 AGGGGAAACTTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 58883
Qy 2061 TACAAATGGGAAACATTTAGAACTTATGAGCTTATGAGCTTATGAGCTTATGAGCTTATG 2120
Db 58882 TACAAATGGGAAACATTTAGAACTTATGAGCTTATGAGCTTATGAGCTTATGAGCTTATG 58823
Qy 2121 ATTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGCTGAGTGTGGAAATCTTCTG 2180
Db 58822 ACTAGCTCCAGATAGCCAGCAATATGATCTCAGAGCTGCTGAGTGTGGAAATCTTCTG 58763
Qy 2181 AAAGCCATTACAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2240
Db 58762 AAAGCCATTACAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 58703
Qy 2241 AAATTTGATGAATGTAAGCAGCTTTTACATTCACCAAAAGGAGGATCACTCAGAGC 2300
Db 58702 AAATTTGATGAATGTAAGCAGCTTTTACATTCACCAAAAGGAGGATCACTCAGAGC 58643
Qy 2301 ATGTGGGAGGT 2312
Db 58642 GTGTGGGAGGT 58631

RESULT 11
AC095988/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-29P9, *** SEQUENCING IN PROGRESS ***
AC095988
AC095988
AC095988.7 GI:30522870
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 258787)
Munzky,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,B.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
```

```
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,C.,
Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Giang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawinsy,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muniz,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoleleh,O., Okwuonu,G., Olarunpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puafo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.D.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,A., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 258787)
Worley,K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258787)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24818127.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEC
Center clone name: CH230-29P9
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 248170 bases at least Q40
```



```

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 188149 bases at least Q40
Consensus quality: 190201 bases at least Q30
Consensus quality: 191458 bases at least Q20
Estimated insert size: 194406; sum-of-contrigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contrigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft,' sequence. It currently
  * consists of 2 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
*
* 1 198699: contig of 198699 bp in length
* 198700 198799: gap of unknown length
* 198800 201330: contig of 2531 bp in length.
FEATURES
      source
        1..201330
          /organism="Rattus norvegicus"
          /mol_type="genomic DNA"
          /db_xref="taxon:10116"
          /clone="CH230-324B23"
      misc_feature
        1..1448
          /note="wgs end_extension
            clone_end:Sp6"
            1920..4036
      misc_feature
        /note="wgs end_extension
            clone_end:Sp6"
            complement(6631..7552)
      misc_feature
        /note="clone boundary
            clone_end:Sp6
            site:
            end_sequence:B2238666"
ORIGIN
Query Match      26.2%; Score 1265.6; DB 2; Length 201330;
Best Local Similarity 77.3%; Pred. No. 1.2e-301;
Matches 1632; Conservative 0; Mismatches 398; Indels 82; Gaps 6;
QY 201 AGATGAACAGTTCTCTACTTTATTCGATACAAAGTGAATAGATTCTAGATCCCC 260
DB 35153 AGATGAACAGTTCTCTACTTTATTCGATACAAAGTGAATAGATTCTAGATCCCTC 35094
QY 261 TTCTCACCGGTGTCATTAAACCCCTACTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 320
DB 35093 TTCTCATCCATTGTCTTTAAACCCCTGCATTAGCCAAATGTGAATGAACCTACAATTTCTAA 35034
QY 321 TGAGTCACAAACACAACTGAAGTCTTCCCTGGCTCATTCAGCTCCCTGACCACAGA 380
DB 35033 TGAGCAGGACACAGCGGAAATCTTTCTTCTGCTGGCTCAGCTCCCTTTACAAGGA 34974
QY 381 GGAAGAGGATCACTGTGCTTAATGACAGAGTGAATCTAAATCCAGAGATTGCCACAAT 440
DB 34973 GGGAAAGGTCCCTGTGCTAACGGACAGACGGTAGTCTAAATCCAGAGATTGACACGAT 34914
QY 441 GTGGATTGATGAATGCTGTGTCAGAGACAGTGAATTAAGAGAACTATAGTTGGGA 500
DB 34913 GTGGATTGATGAATGCTGTGTCAGAGACAGTGAATTAAGAGAACTATACATCAGGA 34854
QY 501 TGATCAATGCACTGTGTTGAAGTGGGAGAGAAATGTGGAACCTGGCTGTCTGCC 560
DB 34853 TGATCAATTTAGTCCGTTGAAGTGGGAGAGAACTGAGGAGCTCACTGTCTGCC 34794
QY 561 AGATGAGAAAGATGTTCTTGTGTAGCGGTGATGATCAATCTGATTAAGAGCAATPACA 620
DB 34793 AGACGAGAAAGATGTTCTTGTGTGTAGCTGTATGATCAATCTGATTAAGAGCACTGCA 34734
QY 621 AAACGATTACAGGATTGTAATAATATAATAGTCAATCCCTTATGGATGCTTTTAGCTG 680

```

```

DB 34733 AAGTGATTTTCAGGATTGTAAATAACTATAATAGTCAATCCCTCATGGATCTTTTAGCTG 34674
QY 681 TTCACCTGGATAAATAAGACAGACAAACTGATCAATTTAGTTTGTAGTATAAATCAGTCCAC 740
DB 34673 TTCACCTGGATAAATAAGACAGACAAACTGATCAATTTAGTTTGTAGTATAAATCAGTCCAC 34614
QY 741 TGAAGAAGATATGAATTCAGAGAAACAAATGGATCCATTGAATAGACCGAAACACAGAGG 800
DB 34613 AGAAAAGGTATTAACCTCAGAGAAACAAATGGATGCTTTGAATAAACCAAAATCAGAGAG 34554
QY 801 GAGATCTGTTAAACCACTGTGCTCTACTTCACTCTGATCTAGTCTAGCCAGTGTCTGTCCCC 860
DB 34553 GGAATCTCTTTAATCATCTCTGCGCTCCTTCAGCTAACAGTGCAACCAAGTATCTCTCTC 34494
QY 861 TTCACAAATTAAGGATGACGGAAGTAGTAGAGACCCCTCCATCTCTCGGATTAACAG 920
DB 34493 TTCACAAATTAAGGATGACGGAAGTAGTAGAGACCCCTCTACATCTACAGCTACAG 34434
QY 921 TTTAAGGTTGATTCAGTAACTCTCATCCAGGAAACAGATGGATGCTCTGTGTTAAAAA 980
DB 34433 CTTAGCAATTAAT-----ACATCCAGGGAATGGATGGAGGCCCTGCCATTAATAA 34383
QY 981 GCAAGAGACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTTAGGACAGA 1040
DB 34382 ACAAGAGGATTATATGCTGTATGAGACCTTTCTGGCGCAAAACAGCTCTTTTAGGACAGA 34323
QY 1041 TCTAGGAGTCCAAATTCCTTTTCCACATGATGAGGGGATTTTGTATGAAAAAGAGCC 1100
DB 34322 TCTAGGAGTCCAAATTCCTTTTCCACATGAGGGGAGGAGCTTTAAACAAAAACAGAGCC 34263
QY 1101 ASCAGAGGAGACCACTGAAGATCCCTCCGCTCTGTTTACCTTCTCTCAAC 1160
DB 34262 AGCAGAGGAGACAGACAGCTGAAGACCCCTCCATCTGTTTATCTTTGAATCTCAAGCC 34203
QY 1161 AGACATGCCTTAATGGGTCTGGAAGGAATAATGACTGTGAACGGTGTTCAGATTGCTGT 1220
DB 34202 AGATACTCTGCTTTTATGTGGAAGGACAAT---TCCGAACCATCTCAGACTGCTGT 34146
QY 1221 GCCTAATGAAGTTAGGGCTGTGAATAATGAAGTTTATGAACATGAAGAACTCTTTGGCAC 1280
DB 34145 GTCTAGTGAGCTAGAACTGATGGAGATGGAGTAAAGGACTCT----- 34103
QY 1281 TACAGAATTCCTTAATATACAGAGCAATTTCTCTGAATCTCAGGACATGACTAATTTGAA 1340
DB 34102 -----CAGACGACTCACTGGGA 34086
QY 1341 GTTGACTAAACTAAATGAGATGAATGATAGCAAGTAAACGAGAAAGGAAAGTTTCT 1400
DB 34085 CTGAGCAAAAGCTAAATGAGTGGATGATGCCAAGAAAGTGAAGAAATCAAGTGGTTTT 34026
QY 1401 ACAGATTGTGACGCTGAGGACACATAATGTTGATAGTGAGGACAGTGTGTTGGATTGCG 1460
DB 34025 ACAGATCAGCCCAACCTGAGGACCCCTCA-----GAGTGCAGGAGAGTGTGTAGATGCG 33972
QY 1461 AGATGAGGTCTAGATTTTAAAGGAACCTTGCAATTAGTGAAGTGAAGAAATGTGATTCTC 1520
DB 33971 AGATTGAGATCTAGATTCTCAAGGAACCTTGCTTGAATGAAGTGAAGATATGATTCTC 33912
QY 1521 CACTGTATAGACACACCGAGCAAAATTTCTATCTAATGTTGTTGTTGTTCTTATGGAAT 1580
DB 33911 CACTGTATGATGACCGAGCAAAATTTCTCTGTTAATAGTTGTTGTTGTTCTTATGGAAT 33852
QY 1581 GCAAGACCCAGGTGTTCTTTTGTTCGAAAGCTTTTACCTCCAAAGAGATTTCAGTAAC 1640
DB 33851 GCAGAGCCCAATTTGATCTTTTGTTCGAAAGCTTTTACCTCCAAAGAGATTTCAGTAAC 33792
QY 1641 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAATATTTATGAACAGAG 1700
DB 33791 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAATATTTATGAACAGAG 33732
QY 1701 AGGAATGAGGCCACAGAGGAGGTGGACTACTTTTAAACAGCAGCTGGTGACCTTAATCAA 1760

```


Matches 1080; Conservative 0; Mismatches 142; Indels 16; Gaps 8;

QY 3605 TCCTGCGAGAGTGCCTCTTCAATGAAGAGCGAGACTCTCATCTTGTGTGTACAGA 3664
 Db 1 TACTGCGAGAGTGCCTCTTCAATGAAGAGCGAGACTCTCATCTTGTGTGTACAGA 60

QY 3665 ATGATGATGGAATATCAGACCCAGGCTATCAGTATTCACAATCAGCCCGAGAAAGTGA 3724
 Db 61 ACATGATGGAATATCAGACCCAGGCTATCAGTATTCACAATCAGCCCGAGAAAGTGA 120

QY 3725 CTGGTGCAGTTCCTTGTGTTTCAGTGGCGCTCTGAATCTCTTCTGATACCTTGCCA 3784
 Db 121 CTGGGCGCATCTTCTTGTGTTTCAGTGGCGCTCTGAATCTCTTCTGATACCTTGCCA 180

QY 3785 AGTCCAGTATTCGGAAGATGTTGTTATGGTCCAGATTACTGCAGAGAACATGGATTCTCT 3844
 Db 181 AGTCCAGTATTCGGAAGATGTTGTTATGGTCCAGATTACTGCAGAGAACATGGATTCTCT 240

QY 3845 TGAGGAGGCACTGCGAGAGATGAAGGACTTCCATCAGTGTGGGAAGCGGACGGG 3904
 Db 241 TGAGGAGGCACTGCGGAGAGATGAAGGACTTACCATCAGTGTGGGAAGCGGATGCGAG 300

QY 3905 AGGAACCCAGAGACATCCATCCAGTGGTGGTATGATGACAGAAACGTTAGCAGG 3964
 Db 301 AGGAACCCAGAGACATCCATCCAGTGGTGGTATGATGACAGAAACGTTAGCAGG 360

QY 3965 GTGTCTAGTCTATAGATGGGAAGTCCATGGAGCTATTAACAAATGTGAAGATATTC 4024
 Db 361 GTGTCTAGTCTATAGATGGGAAGTCCATGGAGCTATTAACAAATGTGAAGATATTC 420

QY 4025 ATGGATCAGATATTAAGCAATGAAGAAATGAAATCAGATGACAGAGTGTCTTCTAG 4084
 Db 421 ATGGATCAGATATTAAGCAATGAAGAAATGAAATCAGATGACAGAGTGTCTTCTAG 480

QY 4085 AAAAGCATGACAGACATATTCCTCAGTGTCTTCCAGATCAGATGACAGAGTGTCTTCTAG 4144
 Db 481 AAAAGCATGACAGACATATTCCTCAGTGTCTTCCAGATCAGATGACAGAGTGTCTTCTAG 540

QY 4145 ATGTGCGCAAGCTTTTTCCTTGTCTCTCTGCTCCTCAGTGTGAGAGAGATG 4204
 Db 541 ATGTGCGCAAGCTTTTTCCTTGTCTCTCTGCTCCTCAGTGTGAGAGAGATG 600

QY 4205 GAATGACCAATCTGGAGTACGTGTGACATCTGATCAGATCAGATGAGTGTGCTATCAGAG 4264
 Db 601 GAATGACCAATCTGGAGTACGTGTGACATCTGATCAGATCAGATGAGTGTGCTATCAGAG 660

QY 4265 GGAGCAATGGCGAGCCCTTCCCTGCGAGTACATGAATGATCTGGATAGCCCTTGTGTC 4324
 Db 661 GGAGCAATGGCGAGCCCTTCCCTGCGAGTACATGAATGATCTGGATAGCCCTTGTGTC 720

QY 4325 CGGTGATCCATGGAGGCGCTGCGAGCTTGTGAGGCGCCGTTGTCTATGGAATCTATCT 4384
 Db 721 CGGTGATCCATGGAGGCGCTGCGAGCTTGTGAGGCGCCGTTGTCTATGGAATCTATCT 780

QY 4385 TTTATATCTGAAACATCTGTATTAACAGAGAGACTTCAATTTTTCTTGTTCAGACTT 4444
 Db 781 TTTATATCTGAAACATCTGTATTAACAGAGAGACTTCAATTTTTCTTGTTCAGACTT 836

QY 4445 GTTGAACAGAGTATACCCCAATCTATTCGACTTTAAACTGGAATTAAGC-TTTT 4503
 Db 837 GTTGAACAGAGTATACCCCAATCTATTCGACTTTAAACTGGAATTAAGC-TTTT 896

QY 4504 GTTAACATATTATGGGCTGGGAATAGGCTGGAGTGGGGTTTGGGAGACGGGTGGG 4563
 Db 897 GTTAACATATTATGGGCTGGGAATAGGCTGGAGTGGGGTTTGGGAGACGGGTGGG 956

QY 4564 AAAGGTGGTGGGGGACCAATGTTTCCATTAATCTTAAGTCTTCTATGCAATGTCCACA 4623
 Db 957 GAAGGTGGT--AGGGGACAGATGTTCCATGATTTCTAAGTCTTCTATGCAAT--CTCCA 1011

QY 4624 AGAAGATCTGGGAGCTTCTGCTCTGCAACAGAGTATGCTATCTCTGAGCTAATCC 4683
 Db 1012 AGAAGATCTGGGAGCTTCTGCTCTGCAACAGAGTATGCTATCTCTGAGCTAATCC 1069

QY 4684 CTCTGTTACTGTTTAGACAAGATTTCCGCTCCTCTCTCAAGATTTTACTTATGGTCACTG 4743
 Db 1070 CTCTGTTACTGTTTAGACAAGATTTCCGCTCCTCTCTCTTATGA-TTACTTCCGTCATGT 1128

QY 4744 GTCAGCA--AATGCTCAATGGGTACCAACCATCACCAGGGTGGGATGGAGGCGCAGAG 4801
 Db 1129 GTCAGATGAATGCTCATATGGGCACCAACCATCACCAGGTCAATGGG-GGGCAGAAA 1187

QY 4802 GGAATAAATATAAGCATCAAAAAAATAAAAAAATAAAAAA 4839
 Db 1188 GGAATAAATAAGCATCGTCAAAAAAATAAAAAAATAAAAAA 1225

RESULT 14
 AC068614 164706 bp DNA linear PRI 09-JAN-2002
 LOCUS Homo sapiens BAC clone RP11-512N1 from 2, complete sequence.
 DEFINITION AC068614
 ACCESSION AC068614.5 GI:15145605
 VERSION HTG.
 KEYWORDS SOURCE Homo sapiens (human)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 164706)
 Sulston, J.E. and Waterston, R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 2 (bases 1 to 164706)
 Kang, K., Abbott, A., Boyer, E. and Dixon, R.
 The sequence of Homo sapiens BAC clone RP11-512N1
 Unpublished (2001)

REFERENCE
 3 (bases 1 to 164706)
 Waterston, R.H.
 Direct Submission

TITLE
 Submitted (05-MAY-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 4 (bases 1 to 164706)
 Waterston, R.H.
 Direct Submission

TITLE
 Submitted (09-AUG-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 5 (bases 1 to 164706)
 Waterston, R.H.
 Direct Submission

TITLE
 Submitted (09-JAN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Aug 9, 2001 this sequence version replaced gi:14628397.

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

Summary Statistics

Center project name: H_NH0512N01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Taten, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-102G8, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-512N1; actual end is at base position 164706 of RP11-512N1.

Polymorphisms exist between RP11-512N1 and RP11-102G8. Data from AC032014 was used to finish this clone AC068614.

FEATURES

source

1.164706
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

/clone="RP11-512N1"
/clone_lib="RPCI-11"

1..138

/rpt_family="MER1_type"

145..804

/rpt_family="L1"

836..1569

/rpt_family="L1"

1685..1848

/rpt_family="MER1_type"

1868..2044

/rpt_family="MER1_type"

2260..2411

/rpt_family="CR1"

2485..3182

/rpt_family="L2"

3172..3293

/rpt_family="L2"

3298..3488

/rpt_family="L2"

3458..3484

/rpt_family="AT-rich"

3617..3718

/rpt_family="MER1_type"

3719..3886

/rpt_family="Alu"

3887..3946

/rpt_family="MER1_type"

3954..4018

/rpt_family="MIR"

5044..5194

/rpt_family="MIR"

5700..5780

/rpt_family="MER1_type"

6317..6551

/rpt_family="MIR"

6824..6871

/rpt_family="CT-rich"

7157..7256

/rpt_family="L2"

8291..8715

repeat_region
8744..8909
/rpt_family="MaLR"
repeat_region
8933..8971
/rpt_family="L2"
repeat_region
9722..9786
/rpt_family="MIR"
repeat_region
10431..10489
/rpt_family="L2"
repeat_region
10686..10897
/rpt_family="L2"
repeat_region
12174..12338
/rpt_family="MIR"
repeat_region
12430..12505
/rpt_family="L1"
repeat_region
12898..12963
/rpt_family="L2"
repeat_region
13850..13945
/rpt_family="Alu"
repeat_region
14692..15044
/rpt_family="MIR"
repeat_region
15938..16906
/rpt_family="L2"
repeat_region
17595..18092
/rpt_family="L1"
repeat_region
18081..18507
/rpt_family="ACHobo"
repeat_region
18374..18426
/rpt_family="ACHobo"
repeat_region
19734..19816
/rpt_family="AT-rich"
repeat_region
19971..20124
/rpt_family="MIR"
repeat_region
20810..20877
/rpt_family="MER103"
repeat_region
21098..21164
/rpt_family="(TGA)n"
repeat_region
21865..22233
/rpt_family="L2"
repeat_region
22328..22439
/rpt_family="MIR"
repeat_region
22440..22664
/rpt_family="MIR"
repeat_region
22665..22747
/rpt_family="MIR"
repeat_region
23939..24091
/rpt_family="MIR"
repeat_region
24668..24701
/rpt_family="MIR"
repeat_region
24783..25084
/rpt_family="(TTTA)n"
repeat_region
25063..25084
/rpt_family="Alu"
repeat_region
25255..25297
/rpt_family="AT-rich"
repeat_region
26227..27186
/rpt_family="L2"
repeat_region
27187..27284
/rpt_family="ERV1"
repeat_region
27285..27610
/rpt_family="ERV1"
repeat_region
27848..28178
/rpt_family="ERV1"
repeat_region
29215..29249
/rpt_family="L2"
repeat_region
30350..30386
/rpt_family="AT-rich"
repeat_region
32688..32732
/rpt_family="MIR"
repeat_region
33373..33665
/rpt_family="MIR"

Query Match 18.1%; Score 875.4; DB 9; Length 164706;
Best Local Similarity 90.2%; Pred. No. 3.8e-205;

Matches 953; Conservative 0; Mismatches 91; Indels 13; Gaps 1;

```

QY 3796 GTGGAAGATGGTGTATATGTTCCAGATTACTGACAGAAACATGATTCCTTGTAGGAGGCA 3855
Db 117131 GTGGATAAAGAAACTGTGTCCAGATCACTGACAGAGACATGGATTCCTTGTAGGAGGCA 117190
QY 3856 CTGCGAGAGATGAAGACATTCACCATCCTCTGGGAGGCGGAGCGGAGAACCCAG 3915
Db 117191 ATGCGAGAGATGAAGACATTCACCATCCTCTGGGAGGCGGAGCGGAGGATCCAG 117250
QY 3916 GAGCAGATCCATCCAGTGGTGGATGATGACAGAACTTAGCAAGGGTGTCTTAAGT 3975
Db 117251 GAGCAGATCCATCCAGTGGTGGATGACAGAACTTAGCAAGTGTCTTAAGT 117310
QY 3976 CCTATGATGGGAATGATGAGACTATTAACAATGGAAGATATTCATGATGACAA 4035
Db 117311 CCTATAATGAGAATGATGAGACTATTAACAATGGAAGATATTCATGATGACAA 117370
QY 4036 TATAAGCAAAATGGAAAGTAAATCATGATGACAGAGGTGTTTTCTTAGAAA----- 4088
Db 117371 TGTAAAGCAAAATGGAAAGTAAATCATGATGACAGAGGTGTTTTCTTAGAAAACACTAA 117430
QY 4089 -----CGATGACCAAGCAAAATGCTCAGTATCTGACAGATCAAGTAGATTGACTGA 4142
Db 117431 ATGATCCTACAAATGATCATTTAGCAATGATCCTGCAAAATCAATTAGATTGACTAA 117490
QY 4143 GCATGTGCAAGGCTTTTGGCTTCTCTGCTCTGCTCCTCCTGACCTGAACTTCTGAGGAGA 4202
Db 117491 GCATGTGCAAGGCTTTTGGCTTCTCTGCTCTGCTCCTCCTGACCTGAACTTCTGAGGAGA 117550
QY 4203 TGAATGACCAAACTGGGACTACGTGTGACACTTGACTCAGATCAGGTGTGCTATCAAGC 4262
Db 117551 TGAATGACCAAACTGGGACTAAGTATGACACTTGACTCAGATCAGGTGTGCTATCAAGC 117610
QY 4263 AGGAGCAATGCCAGCCCTTCCCTCGAGTACATGAATGATCTGATAGGCTTGGT 4322
Db 117611 AGGAGCAATGCCAGCCCTTCCCTCGAGTATGATGATGATCTGAGACACACTTGGT 117670
QY 4323 GCCGTGTATCCATGAGGGGCTGCGAGCTTAGTGAGGGCCCGTGTGTCATGGAATCTAT 4382
Db 117671 GCCGTGTATCCCGAGGGGCTGTCAGCTAGTGAGGGCCCTGTCATGGAATCTAT 117730
QY 4383 CTTTATATCTGGAACATGCTATAACAGAGAACTTCAATTTTCTGTCAGAC 4442
Db 117731 CTTTATATCTGGAACATGCTATAACAGAGAACTTCAATTTTCTGTCAGAC 117790
QY 4443 TTGTTGCAACAGCACTATACCAATCATTTGCACTTTTAACTTGAAGATTAAGCTTT 4502
Db 117791 TTGTTGCAATGGCAGTATACCAATCATTTGCACTTTTAACTTGAAGATTAAGCTTT 117850
QY 4503 TGTTAACACTATTAATAGGGTGGGAATAGGTTGGAGTGGGGTGTGGGAGCGGTGG 4562
Db 117851 TGTTAACACTGTTAATAGGATGGGAATAGGTTGGAGTGGGGTGTGGAGACTGGTAG 117910
QY 4563 GAAAGGGTGTGGGGGACCGATGTTCCATTAATCTTAAGTCTTCTATGCTATGCTCAAC 4622
Db 117911 GAAAGGGTGTGGGGGACCGATGTTCCATTAATCTTAAGTCTTCTATGCTATGCTCAAC 117970
QY 4623 AAGAGATCTGGGACGCTTCTGTTCTCTGCAACAGATTTATGCTATCTTGGCAGCTATCC 4682
Db 117971 AAGAGATCTGGGACGCTTCTGTTCTCTGCAACAGATTTATGCTATCTTGGCAGCTATCC 118030
QY 4683 CTTCTGTTACTGTTTGAAGAAATTCCTGCTCTCTCAAGATTTACTTATGCTATG 4742
Db 118031 CTTCTGTTACTGTTTGAAGAAATTCCTGCTCTCTCAAGATTTACTTATGCTATG 118090
QY 4743 TGCTCAGAAATGCTCAAAATGGGTACCAACATCAAGGGTGGGATGGGAGGCGAGAGG 4802
Db 118091 TGCTCAGAAATGCTCAAAATGGGTACCAACATCAAGGGTGGGATGGGAGGCGAGAGG 118150
QY 4803 GAAATAAATTAAGCATCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4839
Db 118151 GAAATAAATTAAGCATCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 118187

```

RESULT 15
BC038255
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL

REMARK
COMMENT

BC038255 1144 bp mRNA linear ROD 30-SEP-2003
Mus musculus MAD, mothers against decapentaplegic homolog
(Drosophila) interacting protein, receptor activation anchor, mRNA
(CDNA clone IMAGE:5011916), partial cds.
BC038255
BC038255.1 GI:23458383

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1144)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, I., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ussidi, T.B., Toshiyuki, S.,
Carncini, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abrams, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vallaloon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schneringer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1144)
Strausberg, R.
Direct Submission
Submitted (30-SEP-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Caltech, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Akhtar, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dierich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karling, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantrick, S., Thomas, P.J., Touchman, J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Place: 81 Row: n Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene
prediction.

Location/Qualifiers

FEATURES

source 1. .1144
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5011916"
/tissue type="mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMV.""
/clone_lib="NCI_GCAP_Lu29"
/lab_host="DH10B"
/note="vector: pcwv-sport6"

ORIGIN

Query Match 15.3%; Score 740.8; DB 10; Length 1144;
Best Local Similarity 85.7%; Pred. No. 8.1e-172;
Matches 919; Conservative 0; Mismatches 137; Indels 16; Gaps 8;

QY 3755 CTCTGAATCCTCTCTGTGATACCTTGCCAAAGTCCAGTATTGTGGAAGATGGTGTATGG 3814
Db 10 CGCTGAATCCTCTCTGTGATACCTTGCCAAAGTCCAGTATTGTGGAAGATGGTGTATGG 69
QY 3815 TCCAGATTACTGCAGAGACATGATTCCTTGAGGCGAGGACCTGCGAGAGATGAAGACT 3874
Db 70 TCCAGATCACAGCGGAGAACATGGACTCCITTGAGCGAGGCACTGCGGAGATGAAGACT 129
QY 3875 TCACCATCACCTGTGGGAAGCGGAGGAGAACCCAGGAGCACATCCACATCCAGT 3934
Db 130 TTACCATCACCTGTGGGAAGCGGAGTGCAGAGGACCCCGAGAACAGATCCACATCCAGT 189
QY 3935 GGGTGGATGATGACAGAACGTTTACGAGGTTGCTGTAAGTCTTATAGATGGGAAGTCCA 3994
Db 190 GGGTGGATGATGACAGAACGTTTACGAGGTTGCTGTAAGTCTTATAGATGGGAAGTCTA 249
QY 3995 TGGGACTATTAACAAATGTGAAGTATTCCATGATCAGATATATAAGCAATGCAAAAG 4054
Db 250 TGGAACTTATTAACAAATGTGAAGTATTCCATGGTCAAGTATTAAGCAATGCAAAAG 309
QY 4055 TAATCAGATGACAGAGGTGTTTTCTTAGAAAAAGATGACACAGACACAAATTCGCTCAGTG 4114
Db 310 TCATCAGATGACAGAGGTATTTTTCTTGAAGAACGATGACACACATAAATTCGCTCAGTG 369
QY 4115 ATCTGCGATCAGATGATGATGAGTATGCTGAGCATGTTGCCAAGCTTTTGCCTTCTCTCT 4174
Db 370 ACCCTGCGATCAGACAGACTGACTGAGCATGTCGCCAAGGCAATTCGCTCGCCCTCT 429
QY 4175 GTCCTCACTGAACTCTCTGAAGGAAGATGGAATGACCAAACTGGGACTAGCTGTGACAC 4234
Db 430 GCCCCCACTGAGCTTTTAAAGAGATGGAATGACAAACTGGGACTCGCTGTGACGC 489
QY 4235 TTGACTCAGATCAGTGGCTTATCAAGCAGGAGCAATGGCCAGCCCTTCCTCGCAGT 4294
Db 490 TTGACTCAGATCAGTGGCTTATCAAGCAGGAGCAATGGCCAGCCCTTCCTCGCAGT 549
QY 4295 ACATGAATGATCTGGATAGCGCTTGGTGGCGGTGATCCATGGAGGGGCTGCCAGCTTA 4354
Db 550 ACATGAAGATCTGACAGTGCCTTGGTGGCGGTGATCCATGGAGGGGCTGCCAGCTCA 609
QY 4355 GTGAGGGCCCGTGTCTGATGGAATCATCTTTTATATTCTGGAATAACATCGTATAAACAG 4414
Db 610 GTGAAGGCCCTGTCTGATGGAATCATCTTTTATATTCTGGAATAACATCGCAT----AG 665
QY 4415 AGAAGACTCATTTTTTCTGTTACAGACTGTTGCAACAGCAGTATACCCCAATCATTT 4474
Db 666 ATGAGGACTCCATTGTTTCTGTTACAGCTGTTGAGGAGAGTATATCCAAATCATTT 725
QY 4475 GCACTTTAAACTGGAAGATTAAAGC-TTTTGTAACTATTAAATGGGTGGGAATAGG 4533
Db 726 GCACTTTAAACTGGAAGATTAAAGCTTTTGTAACTATGTCAGTAGGAGGGTTTGGGG 785
QY 4534 GTGGAGTGGGGGTTTGGAGAGCGGTGGGAAGAGGTGTTGGGGGGACCGATGTTCCAT 4593
Db 786 TGGGAATAGGACTTAGGGAGATGGGTGGGGGGTGGT--AGGGGACAGATGTTCCAT 843

Search completed: May 18, 2004, 02:10:13
Job time : 12033.2 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2004, 09:46:50 ; Search time 1114.1 seconds
(without alignments)

18451.753 Million cell updates/sec

Title: us-09-744-167-1

Perfect score: 4839

Sequence: 1 gcatactgaatcagcaggac.....atcaaaaaaaaaaaaaaaaaa 4839

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002s:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4839	100.0	4839	3	Aaz50068 Human Sma
2	1977	40.9	1977	9	Adc64443 Human SAR
3	1830	37.8	1830	5	Aas30091 Human lun
4	1830	37.8	1830	5	Aas30092 Human lun
5	1830	37.8	1830	9	Adb33428 Human nov
6	1830	37.8	1830	9	Adb33429 Human nov
7	1545.8	31.9	4823	3	Aaz50070 Xenopus S
8	699.4	14.5	888	7	Aca56987 Human adi
9	576	11.9	576	7	ACA57333 Human adi
10	570.4	11.8	610	5	Aas29903 Human lun
11	570.4	11.8	610	9	Adb33148 Human nov
12	560.6	11.6	6632	3	Aaz50069 Human Sma
13	559	11.6	5402	6	Abk11709 DNA encod
14	333.2	6.9	2678	3	Aaz50071 Xenopus S
15	318	6.6	332	3	Aac30390 Human sec
16	260.2	5.4	498	8	ACH22774 Human adu
17	245.8	5.1	4068	4	Abi28217 Drosophil
18	220.8	4.6	969	6	Abi89832 Human pol
19	203.4	4.2	475	8	Ach16512 Human adu
20	202	4.2	6784	4	Abi28216 Drosophil
21	160	3.3	164	3	Aac07133 Human sec
22	154	3.2	437	7	Abx54300 Bovine ES
23	143.4	3.0	316	9	Adb56467 Toxicity

24	118	2.4	118	5	AAS30093	Human lun
25	118	2.4	118	9	ADB33430	Human nov
26	89.2	1.8	391	8	ACH20104	Human adu
27	63.8	1.3	701	6	ABQ57538	Human col
28	60.8	1.3	289	2	AAV88493	EST clone
29	60	1.2	60	6	ABN41947	Human spl
30	54.8	1.1	1639	5	AAS67754	DNA encod
31	53	1.1	396	3	AAH31021	Human col
32	53	1.1	472	3	AACT7594	Human ORF
33	53	1.1	479	6	ABO99035	Human ORF
34	53	1.1	489	6	ABSG6469	Novel mur
35	53	1.1	731	9	ADC32237	Human nov
36	53	1.1	977	9	ADC30462	Human nov
37	53	1.1	1362	6	AAS17982	Human CBN
38	53	1.1	1434	5	AAH23820	Human tra
39	52.4	1.1	1296	6	ABK49898	Human CBN
40	52.4	1.1	1821	4	AAH15014	Human CBN
41	52.4	1.1	3610	6	ABNS9933	Novel hum
42	51.6	1.1	2000	7	ADA71398	Rice gene
43	50	1.0	476	8	ACH30322	Human tes
44	50	1.0	2138	2	AAx22280	Human tes
45	50	1.0	2138	8	ACD40524	cDNA sequ

ALIGNMENTS

RESULT 1

AAZ50068

ID AAZ50068 standard; cDNA; 4839 BP.

XX AAZ50068;

XX 04-MAY-2000 (first entry)

XX Human Smad Anchor for Receptor Activation protein-1 encoding cDNA.

XX Smad Anchor for Receptor Activation protein; hSARA1; human;

KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;

KW BMP; activin; anti-inflammatory; cytotatic; antiarthritic; vulnary;

KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;

KW inflammatory response; tumour progression; cell proliferation; fibrosis;

KW fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

PH CDS 439..4410

FT /*tag= a

FT /product= "Human SARA-1 protein"

FT /note= "Binds to receptor regulated Smad proteins"

FT /transl_except= (pos:4177..4179, aa:Thr)

FT /transl_except= (pos:4180..4182, aa:Gln)

FT /transl_except= (pos:4198..4200, aa:Gly)

FT /transl_except= (pos:4279..4281, aa:His)

FT /transl_except= (pos:4306..4308, aa:Phe)

FT /transl_except= (pos:4315..4317, aa:Asp)

FT /transl_except= (pos:4324..4326, aa:Lys)

FT /transl_except= (pos:4327..4329, aa:Met)

XX WC200005360-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WC-CA000656.

XX 20-JUL-1998; 98CA-02237701.

XX 10-DEC-1998; 98CA-02253647.

XX (HSCR-) HSC RES & DEV LP.

XX Wrana JL;

XX

DR WPI: 2000-182691/16.
 DR P-PSDB; AAY44749.
 XX
 PT New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 PT inflammation, tumors, fibrogenesis or tissue damage.
 PT
 XX
 PS Claim 7; Page 50-51; 93pp; English.
 XX
 CC The present sequence is the cDNA encoding the human Smad Anchor for
 CC Receptor Activation protein-1 (hSARA1), a TGF-beta modulator. It is
 CC isolated from lambda ZAP human foetal brain cDNA library and has 62%
 CC identity to XSARA1. The hSARA1 proteins bind to receptor-regulated Smad
 CC proteins and ensures its appropriate localisation for activation by a
 CC Type I receptor of a TGF-beta, activin or bone morphogenetic protein
 CC (BMP) signalling pathway. SARA proteins has anti-inflammatory,
 CC cytoskeletal, anti-arthritis and vulnerability activity. The SARA proteins are
 CC useful for developing agents for prevention or treatment of disorders
 CC involving TGF-beta superfamily member signalling pathway. Such agents can
 CC be used for wound healing, scarring, arthritis and fibrosis (e.g., liver
 CC and kidney), in modulating inflammatory and immune responses, tumour
 CC progression, cell proliferation or fibrogenesis and in tissue
 CC morphogenesis. It is also used for protecting, restoring and regenerating
 CC tissues after tissue damage
 CC
 XX Sequence 4839 BP; 1460 A; 965 C; 1140 G; 1274 T; 0 U; 0 Other;
 SQ

Query Match
 Best Local Similarity 100.0%; Score 4839; DB 3; Length 4839;
 Matches 4839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATACCTGAATCAGCAGGACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCCCGAGA 60
 DB 1 GCATACCTGAATCAGCAGGACTGGCTGGTGGTGCAGCAGACATCATGAGTAAGCCCGAGA 60

QY 61 AGTCGTCTCTATCAGCTGTGTAGGGGAAAGGTTTAAACAGTCTCTTAGTGGTG 120
 DB 61 AGTCGTCTCTATCAGCTGTGTAGGGGAAAGGTTTAAACAGTCTCTTAGTGGTG 120

QY 121 TTTCTCTCAGGATGAGAAATTAATCTCAAGCAGAGCTTACAACTGGGCAAGGTGTA 180
 DB 121 TTTCTCTCAGGATGAGAAATTAATCTCAAGCAGAGCTTACAACTGGGCAAGGTGTA 180

QY 181 GATGAATTTGAACAAACGAGATGAACAGTTCTTCTACTTTATTTGGATACAAAGTGG 240
 DB 181 GATGAATTTGAACAAACGAGATGAACAGTTCTTCTACTTTATTTGGATACAAAGTGG 240

QY 241 AATAAGATCTAGATCCCTCTTCAACGGCTGTCATTTAACCTTCTGCGCTCAT 300
 DB 241 AATAAGATCTAGATCCCTCTTCAACGGCTGTCATTTAACCTTCTGCGCTCAT 300

QY 301 AATGAATCTGCAAGTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGCTCAT 360
 DB 301 AATGAATCTGCAAGTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGCTCAT 360

QY 361 TCAGTCCCTTACACAGAGAGAGGATCACTGCTGTAATGGCAGGACTGTAACTTA 420
 DB 361 TCAGTCCCTTACACAGAGAGAGGATCACTGCTGTAATGGCAGGACTGTAACTTA 420

QY 421 AATCCAGAGATGGCCAAATGTGGATTGATGAAATGCTGTGGAGAGACAGTTAAT 480
 DB 421 AATCCAGAGATGGCCAAATGTGGATTGATGAAATGCTGTGGAGAGACAGTTAAT 480

QY 481 AAGAGAACTATAGTTGGATGATCAATGCAAGTGTGTTGAAGTGGAGAGAGAAATGT 540
 DB 481 AAGAGAACTATAGTTGGATGATCAATGCAAGTGTGTTGAAGTGGAGAGAGAAATGT 540

QY 541 GGAAACCTGGCTGTGTCAGATCAGAGAGATGTTCTTTGTTAGCCGTATGCAATAAC 600
 DB 541 GGAAACCTGGCTGTGTCAGATCAGAGAGATGTTCTTTGTTAGCCGTATGCAATAAC 600

QY 601 TGTGATAAAGGACATTACAAACGAGTTTACAGGATTTGTAATTAATTAATAGTCAATCC 660
 DB 601 TGTGATAAAGGACATTACAAACGAGTTTACAGGATTTGTAATTAATTAATAGTCAATCC 660

601 TGTGATAAAGGACATTACAAACGAGTTTACAGGATTTGTAATTAATTAATAGTCAATCC 660
 661 CTTATGATGCTTTTACGCTGTTTACCTGGATTAATGAAAAACAGACAACTGATCAATTTAGT 720
 720 CTTATGATGCTTTTACGCTGTTTACCTGGATTAATGAAAAACAGACAACTGATCAATTTAGT 720
 721 TTTAGTATAAATGAGTCCACTGAAAAAGATATGAATTCAGAGAAACAAATGATCATCTG 780
 780 TTTAGTATAAATGAGTCCACTGAAAAAGATATGAATTCAGAGAAACAAATGATCATCTG 780
 781 AATAGACCGAAAAACAGAGGGGAGATCTGTTAAACCATCTCTGCTCTCTATCTGATAGT 840
 840 AATAGACCGAAAAACAGAGGGGAGATCTGTTAAACCATCTCTGCTCTCTATCTGATAGT 840
 841 CTAGCCAGTGTCTGTTCCCTTCCCAATTAAGGATGACGGAAGTATAGTAGAGACCC 900
 900 CTAGCCAGTGTCTGTTCCCTTCCCAATTAAGGATGACGGAAGTATAGTAGAGACCC 900
 901 TCCATGCTCGCAATTACAAGTTTAAACGTTGATTCAAGTAACTCTCATCCAGGAAACAGAT 960
 960 TCCATGCTCGCAATTACAAGTTTAAACGTTGATTCAAGTAACTCTCATCCAGGAAACAGAT 960
 961 GGATGCTCTGCTTTAAAAAGCAGAGAACTATATACAGATGAGGACCTCTACTGCGAAA 1020
 1020 GGATGCTCTGCTTTAAAAAGCAGAGAACTATATACAGATGAGGACCTCTACTGCGAAA 1020
 1021 ATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTTCCACATGATGAGGAGG 1080
 1080 ATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTTCCACATGATGAGGAGG 1080
 1081 ATTTGATGAAAAAGAGCCAGCAGAGAGAGACCACTGAAAGAAATCCCTCCGGTCTGCT 1140
 1140 ATTTGATGAAAAAGAGCCAGCAGAGAGAGACCACTGAAAGAAATCCCTCCGGTCTGCT 1140
 1141 TTACCTTTGCTTCTCAAAACAGACATGCTTAATGGTCTGGAAGGAATTAATGACTGTGAA 1200
 1200 TTACCTTTGCTTCTCAAAACAGACATGCTTAATGGTCTGGAAGGAATTAATGACTGTGAA 1200
 1201 CGGTGTCAGATGCTTGTGCTTAATGAAAGTTAGGGCTGATGAAATGAAAGTTATGAA 1260
 1260 CGGTGTCAGATGCTTGTGCTTAATGAAAGTTAGGGCTGATGAAATGAAAGTTATGAA 1260
 1261 CATGAAGAACTCTTGGCCTACAGATTCCTTAATGACAGAGCATTTCTCTGAATCT 1320
 1320 CATGAAGAACTCTTGGCCTACAGATTCCTTAATGACAGAGCATTTCTCTGAATCT 1320
 1321 CAGGACATGACTAATTTGGAAGTTGACTAAACTAAATGAGATGAATGATGACCAAGTAAAC 1380
 1380 CAGGACATGACTAATTTGGAAGTTGACTAAACTAAATGAGATGAATGATGACCAAGTAAAC 1380
 1381 GAAGAAAGAAAGTTTCTACAGATGATGACCTGAGGAGACACTAATGGTATAGTGA 1440
 1440 GAAGAAAGAAAGTTTCTACAGATGATGACCTGAGGAGACACTAATGGTATAGTGA 1440
 1441 GGACAGTGTGTTGGATTGGCAGATGAGATTTAAAGGAACTTGCATTAAGTGA 1500
 1500 GGACAGTGTGTTGGATTGGCAGATGAGATTTAAAGGAACTTGCATTAAGTGA 1500
 1501 AGTGAAGAAATGTGATTTCTCCACTGTTATAGACACACAGAGGAGAAATTAATCTATTAAT 1560
 1560 AGTGAAGAAATGTGATTTCTCCACTGTTATAGACACACAGAGGAGAAATTAATCTATTAAT 1560
 1561 GGTGTTGATTTCTATGGAATGCAAGCCAGGTTCTTTTGTTCCTCAAGACTTTTACCC 1620
 1620 GGTGTTGATTTCTATGGAATGCAAGCCAGGTTCTTTTGTTCCTCAAGACTTTTACCC 1620
 1621 TCCAAAGAAAGTTTCAAGTAAACAGAGAAAGAAATAGAGGAAAGCAAGTCAAGATGCTAC 1680
 1680 TCCAAAGAAAGTTTCAAGTAAACAGAGAAAGAAATAGAGGAAAGCAAGTCAAGATGCTAC 1680
 1681 TCAATATTTATGACAGAGAGGAAATGAGGCCACAGAGAGGAGTGGACTACTTTTAAAC 1740
 1740 TCAATATTTATGACAGAGAGGAAATGAGGCCACAGAGAGGAGTGGACTACTTTTAAAC 1740

QY 1741 AGCACTGGTGACCTAATGAAGAAAAATTAATTAATAAATTCGTAGTCAAGTTCATCA 1800
DB 1741 AGCACTGGTGACCTAATGAAGAAAAATTAATTAATAAATTCGTAGTCAAGTTCATCA 1800
QY 1801 GTGCTGGGCAATCTCCCAAGAGTAGTAGAGCGCTGCCATCTACGTGTCTCTTTT 1860
DB 1801 GTGCTGGGCAATCTCCCAAGAGTAGTAGAGCGCTGCCATCTACGTGTCTCTTTT 1860
QY 1861 GGTGGTGCAGAGCCCAAGCAACCTCTAAATCTTAAACTTCAAAATCCAAAGCCATTATCA 1920
DB 1861 GGTGGTGCAGAGCCCAAGCAACCTCTAAATCTTAAACTTCAAAATCCAAAGCCATTATCA 1920
QY 1921 GACCAATTAATAATGACTTCTCGTGCAGAGTGGAAATTAATTAATAAATAAATATGAT 1980
DB 1921 GACCAATTAATAATGACTTCTCGTGCAGAGTGGAAATTAATTAATAAATAAATATGAT 1980
QY 1981 ATTCTTGGGAAGCAAAATTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTG 2040
DB 1981 ATTCTTGGGAAGCAAAATTAGGGGAAACTCAGCAACCAATGTATGAGTCCATCTTG 2040
QY 2041 GGAACATCTCTAATGTCGATCAAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
DB 2041 GGAACATCTCTAATGTCGATCAAAATGGGAAACATTTAGAAAGTTATGAGGCTGAGATC 2100
QY 2101 TCCACTAGACCATGCTTGCATTTAGCTCCAGATAGCCAGATTAATGATCTCAGAGCTGGT 2160
DB 2101 TCCACTAGACCATGCTTGCATTTAGCTCCAGATAGCCAGATTAATGATCTCAGAGCTGGT 2160
QY 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACCAACGCTGGGTGAGTGGTCCAGTATGG 2220
DB 2161 CAGTTTGGAAATTTCTGCCAGAAAGCCATTACCAACGCTGGGTGAGTGGTCCAGTATGG 2220
QY 2221 GTACCGGATTTCTCAGGCTCAGAAATGTGATGAATGTGAAGCCAGGTTTACATTCACCAA 2280
DB 2221 GTACCGGATTTCTCAGGCTCAGAAATGTGATGAATGTGAAGCCAGGTTTACATTCACCAA 2280
QY 2281 AGGAGGATCAGCTCAGAGATGTGGAAAGTTTCTGCTCTCTGCTAGCTGAA 2340
DB 2281 AGGAGGATCAGCTCAGAGATGTGGAAAGTTTCTGCTCTCTGCTAGCTGAA 2340
QY 2341 TGTAAATCTGTTATACATGGAAGAAAGAGCTAGAGTGTGTAAATCTGCCATTCAGTG 2400
DB 2341 TGTAAATCTGTTATACATGGAAGAAAGAGCTAGAGTGTGTAAATCTGCCATTCAGTG 2400
QY 2401 CTAAATGATGCTCAAGCTGGAGAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCT 2460
DB 2401 CTAAATGATGCTCAAGCTGGAGAACATGATGAGTGCCTCAAGCCAGAGCCCTTAACCT 2460
QY 2461 AACAACTCTGCTGAATACTGTTCTACTATCCCTCTTGCAAGAGCTCAGGCTCAGGA 2520
DB 2461 AACAACTCTGCTGAATACTGTTCTACTATCCCTCTTGCAAGAGCTCAGGCTCAGGA 2520
QY 2521 GCTCTGAGCTCTCACCTCCACTGTGATGTACTCTGGAGTTTAAAGCCCTTGA 2580
DB 2521 GCTCTGAGCTCTCACCTCCACTGTGATGTACTCTGGAGTTTAAAGCCCTTGA 2580
QY 2581 GCAGAAGTGGTTCAGCCAGAGAGCAGAGCGAGTTTGGTTGTGATGGGATCTTGCCC 2640
DB 2581 GCAGAAGTGGTTCAGCCAGAGAGCAGAGCGAGTTTGGTTGTGATGGGATCTTGCCC 2640
QY 2641 AATGGAGAAGTTGCTGATGAGCCAAATTAACAATGAATGAAATTCCTCTGAGGAACC 2700
DB 2641 AATGGAGAAGTTGCTGATGAGCCAAATTAACAATGAATGAAATTCCTCTGAGGAACC 2700
QY 2701 CTGCTGTGTACAGGACCCAGTCAAGCCAGTACTACCTCTACAGCAGAGACG 2760
DB 2701 CTGCTGTGTACAGGACCCAGTCAAGCCAGTACTACCTCTACAGCAGAGACG 2760
QY 2761 GATATTTCTATTTCTCTGGAGTATTAACCTCAGGTGGAAATTCCTGTTGGAAGTGCATG 2820
DB 2761 GATATTTCTATTTCTCTGGAGTATTAACCTCAGGTGGAAATTCCTGTTGGAAGTGCATG 2820

QY 2821 AATCTTATTCCTGAAGATGGCTTCTCCCAATTCCTCATCTCCACTGGTGTAAAGGAC 2880
DB 2821 AATCTTATTCCTGAAGATGGCTTCTCCCAATTCCTCATCTCCACTGGTGTAAAGGAC 2880
QY 2881 TATGCTGTGGAAGAGAAACCATCACAGATTTTCAAGTAATGACAGAGTTGAGGATGGTGGC 2940
DB 2881 TATGCTGTGGAAGAGAAACCATCACAGATTTTCAAGTAATGACAGAGTTGAGGATGGTGGC 2940
QY 2941 CCTGACCCACTGTATTTTGTGTTTAAATGCAAAATTTGTTGTAATGTTTAAATTTGTAAT 3000
DB 2941 CCTGACCCACTGTATTTTGTGTTTAAATGCAAAATTTGTTGTAATGTTTAAATTTGTAAT 3000
QY 3001 TATGTGAACAGAGTGTGTTTCAACCAAGGAAATGTCATGAGTGGTGGTGGT 3060
DB 3001 TATGTGAACAGAGTGTGTTTCAACCAAGGAAATGTCATGAGTGGTGGTGGT 3060
QY 3061 GAGATAGTCAITCTTCTACAGTGTTTTACCGGATGAAAGTGTGTTGCCAAAGATATCTTT 3120
DB 3061 GAGATAGTCAITCTTCTACAGTGTTTTACCGGATGAAAGTGTGTTGCCAAAGATATCTTT 3120
QY 3121 AATCACTTTGTCAGCTTTATCGGATGCTCTGGCAGGAAATGTTGTCAGCAACTGGGA 3180
DB 3121 AATCACTTTGTCAGCTTTATCGGATGCTCTGGCAGGAAATGTTGTCAGCAACTGGGA 3180
QY 3181 CATTCCTTCTCAGTCAAAAGTTTCTTGGCAGTAAAGAACATGTTGATGATG 3240
DB 3181 CATTCCTTCTCAGTCAAAAGTTTCTTGGCAGTAAAGAACATGTTGATGATG 3240
QY 3241 ACATCTACCTACAGTCAAGACCTTAGTACTCCCAACCCACCTTACTTGTGGG 3300
DB 3241 ACATCTACCTACAGTCAAGACCTTAGTACTCCCAACCCACCTTACTTGTGGG 3300
QY 3301 ATTCTTATCCAAATGGAAACTCTCTGGGCTAAAGTATTTCCCTATCCGTCTGATGTTG 3360
DB 3301 ATTCTTATCCAAATGGAAACTCTCTGGGCTAAAGTATTTCCCTATCCGTCTGATGTTG 3360
QY 3361 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTGAGTTTGGAG 3420
DB 3361 AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTGAGTTTGGAG 3420
QY 3421 CCATTTGTTGGAGAGCGGGCATACCATCATGATTTCTTGGCAGCTTCAGAAATATAC 3480
DB 3421 CCATTTGTTGGAGAGCGGGCATACCATCATGATTTCTTGGCAGCTTCAGAAATATAC 3480
QY 3481 CAGTATACCTTCCCAAGTGTTCAGAGTTTGGTGTGATATGGAAGTTTGGAAAACTAGC 3540
DB 3481 CAGTATACCTTCCCAAGTGTTCAGAGTTTGGTGTGATATGGAAGTTTGGAAAACTAGC 3540
QY 3541 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAGCCATGAACAAGTCCCAATGAG 3600
DB 3541 ATCAAAATTTCCAGCAACAGATACATGAGATGATGAAGCCATGAACAAGTCCCAATGAG 3600
QY 3601 CATGCTCTGGCAGGAGTGCCTGCTTCAATGAAAGGCGAGCTCTCATCTTGTGTGTA 3660
DB 3601 CATGCTCTGGCAGGAGTGCCTGCTTCAATGAAAGGCGAGCTCTCATCTTGTGTGTA 3660
QY 3661 CAGAAATGATGAGAAACTATCAGACCCAGGCTATCAGTATTCACAAATCAGCCCCAGAAA 3720
DB 3661 CAGAAATGATGAGAAACTATCAGACCCAGGCTATCAGTATTCACAAATCAGCCCCAGAAA 3720
QY 3721 GTGACTGTGTCAGTTTCTTGTGTTTCTGAGGCTCTGAAATCTCTTCTGGATACCTT 3780
DB 3721 GTGACTGTGTCAGTTTCTTGTGTTTCTGAGGCTCTGAAATCTCTTCTGGATACCTT 3780
QY 3781 GCCAAGTCCAGTATTTGTGAAGATGGTGTATGTTGTCAGATTAATCTGAGAGAACATGGAT 3840
DB 3781 GCCAAGTCCAGTATTTGTGAAGATGGTGTATGTTGTCAGATTAATCTGAGAGAACATGGAT 3840
QY 3841 TCCCTGAGCAGGCACTGCGAGAGATGAGGACTTCAACATCAGCTGTGGAGCGGAC 3900
DB 3841 TCCCTGAGCAGGCACTGCGAGAGATGAGGACTTCAACATCAGCTGTGGAGCGGAC 3900
QY 3901 GCGGAGGAACCCAGGAGCACATCCACATCCAGTGGGTGGATGATGACAAAGAACGTTAGC 3960

```

Db 3901 GGGAGGAGACCCAGGAGCATCCATCCAGTGGGTGGAGATGACAAAGACGTTAGC 3960
Qy 3961 AAGGCTGTCGTAAGTCCATATAGATGGGAAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
Db 3961 AAGGCTGTCGTAAGTCCATATAGATGGGAAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
Qy 4021 TTCATGATCAGAAATATAAGCAAAATGAAAAGTAATCAGATGAGACAGAGGTGTTTTTC 4080
Db 4021 TTCATGATCAGAAATATAAGCAAAATGAAAAGTAATCAGATGAGACAGAGGTGTTTTTC 4080
Qy 4081 CTAGAAAACGATGACACAGCAAAATGCGCTCAGTGCATCTGCAGATCAGATGAGATGACT 4140
Db 4081 CTAGAAAACGATGACACAGCAAAATGCGCTCAGTGCATCTGCAGATCAGATGAGATGACT 4140
Qy 4141 GAGCATGTTGCCAAAGCTTTTGGCTTGCTCTCTGTCTCCTACCTGAAACTTCTGAAGAA 4200
Db 4141 GAGCATGTTGCCAAAGCTTTTGGCTTGCTCTCTGTCTCCTACCTGAAACTTCTGAAGAA 4200
Qy 4201 GATGAATGACCAACTGGGACTAGTGTGACATCTGACTCAGATCAGATGAGTGGCTATCAA 4260
Db 4201 GATGAATGACCAACTGGGACTAGTGTGACATCTGACTCAGATCAGATGAGTGGCTATCAA 4260
Qy 4261 GCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGATGATCTGGATAGCGCCTTG 4320
Db 4261 GCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGATGATCTGGATAGCGCCTTG 4320
Qy 4321 GTGCGGTGATCCATGGAGGGGCTGCCAGCTTAGTGAGGGCCCGTGTGATGGAATC 4380
Db 4321 GTGCGGTGATCCATGGAGGGGCTGCCAGCTTAGTGAGGGCCCGTGTGATGGAATC 4380
Qy 4381 ATCTTTTATTTCTGAAAACATCGTATAAAGAGAGAGCTTCATTTTTCGTTTCAG 4440
Db 4381 ATCTTTTATTTCTGAAAACATCGTATAAAGAGAGAGCTTCATTTTTCGTTTCAG 4440
Qy 4441 ACTTGTTCGACAGCAGTCATACCAAAATCAATTTGCACTTTAAACTGGAAGATTAAGCT 4500
Db 4441 ACTTGTTCGACAGCAGTCATACCAAAATCAATTTGCACTTTAAACTGGAAGATTAAGCT 4500
Qy 4501 TTTGTTTACATTAATTAAGGGTGGGAATAGGTGGGAGTGGGGTTTGGGAGACGGT 4560
Db 4501 TTTGTTTACATTAATTAAGGGTGGGAATAGGTGGGAGTGGGGTTTGGGAGACGGT 4560
Qy 4561 GGGAAAGGGTGGTGGGGGACCGATGTTCCATAATTTCAAGTCTTCTATGCTATGTTCCA 4620
Db 4561 GGGAAAGGGTGGTGGGGGACCGATGTTCCATAATTTCAAGTCTTCTATGCTATGTTCCA 4620
Qy 4621 CCAAGAGATCGGCAGCTTCTGTTCTCTGACACAGATGATGCTATCTCTGAGCTAAT 4680
Db 4621 CCAAGAGATCGGCAGCTTCTGTTCTCTGACACAGATGATGCTATCTCTGAGCTAAT 4680
Qy 4681 CCCCTTCTGTTACTGTTTAGACAAGAAATTCGGCTCTCTCTCAAGATTTACTTATGTTCA 4740
Db 4681 CCCCTTCTGTTACTGTTTAGACAAGAAATTCGGCTCTCTCTCAAGATTTACTTATGTTCA 4740
Qy 4741 TGTGCTCAGAAATGCTCAATGGGTAGCAACCATCACCAGGGTGGGATGGAGGGGAGAG 4800
Db 4741 TGTGCTCAGAAATGCTCAATGGGTAGCAACCATCACCAGGGTGGGATGGAGGGGAGAG 4800
Qy 4801 GGGAAATATAAATATAAGCATCAAAAAAATAAAAAA 4839
Db 4801 GGGAAATATAAATATAAGCATCAAAAAAATAAAAAA 4839

```

RESULT 2

ADC6443

ID ADC6443 standard; DNA; 1977 BP.

XX AC

XX AC ADC6443;

XX DT

XX DT 18-DEC-2003 (first entry)

XX DE

DE Human SARA_v5 coding sequence.

```

XX Hepatotropic; vaccine; protein-protein interaction;
XW Transforming Growth Factor beta; TGF beta; hepatitis;
XW Selected Interacting Domain; SID; bait; human; SARA_v5; gene; ds.
XX Homo sapiens.
OS
XX WO2003045990-A2.
XX 05-JUN-2003.
XX 26-NOV-2002; 2002WO-EP013866.
XX 26-NOV-2001; 2001US-033348P.
PR 31-MAY-2002; 2002US-0384537P.
PR 30-OCT-2002; 2002US-0422471P.
XX (HYBR-) HYBRIGENICS.
XX Legrain P, Gauthier J, Colland F, Jacq X;
XX WPI; 2003-505185/47.
XX P-PSDB; ADC64239.
XX New complex between two interacting proteins, useful for screening
PT molecules that inhibit transforming growth factor beta (TGF beta) or TGF
PT beta super-family of cytokines pathway for diagnosing or treating TGF
PT beta diseases or disorders.
XX Disclosure; SEQ ID NO 10; 148pp; English.
XX The present invention relates to protein-protein interactions and
CC complexes involved in Transforming Growth Factor (TGF) beta disorders
CC and/or diseases. The complex between two interacting proteins is useful
CC for screening molecules that inhibit TGF beta for diagnosing or treating
CC diseases or disorders involving TGF beta e.g., hepatitis. To illustrate
CC the invention, Selected Interacting Domains (SID) of proteins and their
CC coding sequences were isolated (ADC6423-ADC64318). Sequences ADC6434-
CC ADC64242 are bait sequences used to isolate the SIDs of the invention.
XX Sequence 1977 BP; 538 A; 433 C; 485 G; 521 T; 0 U; 0 Other;
SQ
Query Match 40.9%; Score 1977; DB 9; Length 1977;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2434 ACTGCTCAGCCAGCAGCCCTAACCTAACATCTCTGTAATCTGTCTACTATCCCT 2493
Db 1 AGTGGCTCAGCCAGCAGCCCTAACCTAACATCTCTGTAATCTGTCTACTATCCCT 60
Qy 2494 CCCTTGCAGCAAGCTCAGGCCCTCAGAGCTCTGAGCTCTCCACCTCCACCTGTGATGTA 2553
Db 61 CCCTTGCAGCAAGCTCAGGCCCTCAGAGCTCTGAGCTCTCCACCTCCACCTGTGATGTA 120
Qy 2554 CTTGTGGAGATTTAAAGACCCCTGGAGCAGAGTGGCTCAGCCAGAGCAGAGGGCGA 2613
Db 121 CTTGTGGAGATTTAAAGACCCCTGGAGCAGAGTGGCTCAGCCAGAGCAGAGGGCGA 180
Qy 2614 GTTTCGTTTGTGATGGGATCTTGCCCAATGGAGAAGTTGCTGATGACGCCAAATTAACA 2673
Db 181 GTTTCGTTTGTGATGGGATCTTGCCCAATGGAGAAGTTGCTGATGACGCCAAATTAACA 240
Qy 2674 ATGAATGGAATCTCTCTGAGGAACCCCTGGCTGTGTCTCAGAGCCAGTCAAGCCAGTA 2733
Db 241 ATGAATGGAATCTCTCTGAGGAACCCCTGGCTGTGTCTCAGAGCCAGTCAAGCCAGTA 300
Qy 2734 ACTACAGTCTCTCTACAGCAGAGACGGATATTTGCTTATTTCTTGGAGTATAACTCAG 2793
Db 301 ACTACAGTCTCTCTACAGCAGAGACGGATATTTGCTTATTTCTTGGAGTATAACTCAG 360
Qy 2794 GTTGAAGTCTCTGTTGGAGTGCATGAATCTTATTTCTGAAGTGGCTTCTCCCAATT 2853
Db 361 GTTGAAGTCTCTGTTGGAGTGCATGAATCTTATTTCTGAAGTGGCTTCTCCCAATT 420

```


disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma and pleurisy. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 37.8%; Score 1830; DB 5; Length 1830;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

201 AGATGAACAGCTTCTCTACTTTATTTGGATACAAAGTGGATPAAGATTTCTAGATCCCC 260
1 AGATGAACAGCTTCTCTACTTTATTTGGATACAAAGTGGATPAAGATTTCTAGATCCCC 60

261 TTCTCAGCGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 320
61 TTCTCAGCGCTGTCAATTAACCCCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTAA 120

321 TGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCATTGAGTCCCTGACACACAGA 380
121 TGAGTCAACACCACTGAAAGTCTTCTCCCTGGCTCATTGAGTCCCTGACACACAGA 180

381 GGAAGAGGATCACTGTCTAATGAGCAGAGCTGTAAATCTAAATCCAGAGATTCGCCAAT 440
181 GGAAGAGGATCACTGTCTAATGAGCAGAGCTGTAAATCTAAATCCAGAGATTCGCCAAT 240

441 GTGAGTGTGAAAGTCTCTTCAGAGACAGCTGTAAATTAAGAGAACTATAGTTGGGA 500
241 GTGAGTGTGAAAGTCTCTTCAGAGACAGCTGTAAATTAAGAGAACTATAGTTGGGA 300

501 TGATCAATGCGAGTCTGTGAAGTGGAGAGAAATGTGAAACCTGGCTTGTCTGCC 560
301 TGATCAATGCGAGTCTGTGAAGTGGAGAGAAATGTGAAACCTGGCTTGTCTGCC 360

561 AGATGAGAAGATGTTCTTGTGTAGCCGTCATGATTAAGTGAATAAGAGCATTACA 620
361 AGATGAGAAGATGTTCTTGTGTAGCCGTCATGATTAAGTGAATAAGAGCATTACA 420

621 AAACGATTTACAGGATTTAATAATTAATAGTCAATCCCTTATGGATGCTTTTAGCTG 680
421 AAACGATTTACAGGATTTAATAATTAATAGTCAATCCCTTATGGATGCTTTTAGCTG 480

681 TTCACTGGATAATGAAGACAGCAAACTGATCAATTTAGTTTATGATTAATGAGTCCAC 740
481 TTCACTGGATAATGAAGACAGCAAACTGATCAATTTAGTTTATGATTAATGAGTCCAC 540

741 TGAABAGATATGAATTCAGAGAACAAATGGATTCATTAAGTGAATAAGAGCAGAGG 800
541 TGAABAGATATGAATTCAGAGAACAAATGGATTCATTAAGTGAATAAGAGCAGAGG 600

801 GAGATCTGTTAAACCATCTGTCTACTTTCATCTGATGATGCTAGCCAGTGTCTGTCCCC 860
601 GAGATCTGTTAAACCATCTGTCTACTTTCATCTGATGATGCTAGCCAGTGTCTGTCCCC 660

861 TTCAATTAAGATGACGGAATATAGTGAAGCCCTCCATGCTGTCGGATTACAAG 920
661 TTCAATTAAGATGACGGAATATAGTGAAGCCCTCCATGCTGTCGGATTACAAG 720

921 TTTAACGGTTCAGTTCAGTAACTCATCCAGGAAACAGATGGATGTCCTGCTGTTAAAA 980
721 TTTAACGGTTCAGTTCAGTAACTCATCCAGGAAACAGATGGATGTCCTGCTGTTAAAA 780

991 GCAAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCTTAGACAGA 1040
781 GCAAGAGAACTATATACAGATGAGGACCTCAGTGGCAAAATCAGCTCTCTTAGACAGA 840
1041 TCTAGGAGTCCAAATCTCTTTCCCATGAGTGGGATTTGATGAAAAAGAGCC 1100
841 TCTAGGAGTCCAAATCTCTTTCCCATGAGTGGGATTTGATGAAAAAGAGCC 900
1101 AGCAGAGGAGACCCACTGAAGATCCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 1160
901 AGCAGAGGAGACCCACTGAAGATCCCTCGGCTCTGGTTTACCTTTGCTCTCAAAAC 960
1161 AGCAGTGCCTAATGGCTGCGAAGGAATTAAGTCTGTAACGGTTCAGATTCGCTTGT 1220
961 AGCAGTGCCTAATGGCTGCGAAGGAATTAAGTCTGTAACGGTTCAGATTCGCTTGT 1020
1221 GCCTAATGAAGTTAGGCTGATGAAATGAAGGTTATGAACATGAAGAAATCTTGGCAC 1280
1021 GCCTAATGAAGTTAGGCTGATGAAATGAAGGTTATGAACATGAAGAAATCTTGGCAC 1080
1281 TACAGATTCCTTAATATGACAGAGCATTCTCTGAAATCTCAGGACATGATTAATGGAA 1340
1081 TACAGATTCCTTAATATGACAGAGCATTCTCTGAAATCTCAGGACATGATTAATGGAA 1140
1341 GTTGACTAAACTAAATGAGATGAATGATAGCAAGTAAACGAAAGAAAGGATTTCT 1400
1141 GTTGACTAAACTAAATGAGATGAATGATAGCAAGTAAACGAAAGAAAGGATTTCT 1200
1401 ACAGATTAAGTCAAGCTGAGGACACTAATGATGATGAGGAGACAGTGTGTGATGGC 1460
1201 ACAGATTAAGTCAAGCTGAGGACACTAATGATGATGAGGAGAGTGTGTGATGGC 1260
1461 AGATCAGGCTGATGATTTAAAGGAACTTGATAGTGAAGTGAAGTGTGATTTCTC 1520
1261 AGATCAGGCTGATGATTTAAAGGAACTTGATAGTGAAGTGAAGTGTGATTTCTC 1320
1521 CACTGTTATAGACACACAGCAGCAAAATTAATCTAATGTTGTGATTTCTATGGAAT 1580
1321 CACTGTTATAGACACACAGCAGCAAAATTAATCTAATGTTGTGATTTCTATGGAAT 1380
1581 GCAAGACCCAGTGTCTTTTGTTCGAAAGCTTTACCTCCGAAAGAGATTCAGTAAC 1640
1381 GCAAGACCCAGTGTCTTTTGTTCGAAAGCTTTACCTCCGAAAGAGATTCAGTAAC 1440
1641 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATTTTATGAACAG 1700
1441 AGAAGAAAAAGAAATAGAGGAAAGCAAGTCAAGATGCTACTCAAAATTTTATGAACAG 1500
1701 AGAAGTGAAGGCAAGAGGAGTGGACTACTTTTAAACAGCAGCTGTGACCTAATGA 1760
1501 AGAAGTGAAGGCAAGAGGAGTGGACTACTTTTAAACAGCAGCTGTGACCTAATGA 1560
1761 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCCTCATGCTGGCAATCTTCCCC 1820
1561 GAAAAATTTATTTACATAATTTCTGTAGTCAAGTTCCTCATGCTGGCAATCTTCCCC 1620
1821 CAAAGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTTGGTGTGCAAGCCCAAGCA 1880
1621 CAAAGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTTTGGTGTGCAAGCCCAAGCA 1680
1881 ACCTTCTAATCTTAAACTTCAAAATTCGAAAGCCATTATCAGACCATTTTACAAAATGACT 1940
1681 ACCTTCTAATCTTAAACTTCAAAATTCGAAAGCCATTATCAGACCATTTTACAAAATGACT 1740
1941 TCTCGCAACAGTGGAAATTAATCTAAAAATAAAAATGATATTTCTTGGGAAAGCAAAAT 2000
1741 TCTCGCAACAGTGGAAATTAATCTAAAAATAAAAATGATATTTCTTGGGAAAGCAAAAT 1800
2001 AGGGGAAACTCAGCAACCAATGTATGAG 2030
1801 AGGGGAAACTCAGCAACCAATGTATGAG 1830

RESULT 4
 AAS30092
 ID AAS30092 standard; DNA; 1830 BP.
 XX
 AC AAS30092;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human lung antigen genomic DNA #162.
 XX
 KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
 KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
 KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
 KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
 KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;
 KW gastrointestinal disorder; renal disorder; respiratory disorder;
 KW wound healing; skin aging; organ transplantation; food preservative;
 KW tissue regeneration; anti-infertility; food additive.
 XX
 OS Homo sapiens.
 XX
 XX WO200155303-A2.
 XX
 XX 02-AUG-2001.
 PD
 XX 17-JAN-2001; 2001WO-US001301.
 XX
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0198874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216860P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246534P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.

Tue May 18 15:16:58 2004

PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249247P.
PR	17-NOV-2000;	2000US-0249248P.
PR	17-NOV-2000;	2000US-0249249P.
PR	17-NOV-2000;	2000US-0249250P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	08-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-457723/49.	
PX	Isolated polypeptide for treating, preventing and/or prognosing	
PT	respiratory disorders related to the lung including lung cancers and also	
PT	for testing and detection e.g. diagnosis.	
XX	Claim 1; SEQ ID NO 356; 507pp; English.	
PS	Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode	
CC	the lung antigen polypeptides of the invention. Lung antigen polypeptides	
CC	and their associated polynucleotides are useful in the diagnosis,	
CC	treatment and prevention of various types of disorders in e.g. humans,	
CC	mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A	
CC	pathological condition can be determined by detecting the presence or	
CC	absence of a mutation in a lung antigen polynucleotide. The treatable	
CC	disorders include autoimmune diseases such as rheumatoid arthritis,	
CC	hyperproliferative disorders such as cardiac arrest, cerebrovascular	
CC	cardiovascular disorders such as cerebral ischaemia, nervous system disorders such as	
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi,	
CC	ocular disorders such as corneal infection, endocrine disorders such as	
CC	premature labour and infertility, gastrointestinal disorders such as	
CC	Crohn's disease, renal disorders, such as glomerulonephritis and	
CC	respiratory disorders such as asthma and pleurisy. The polypeptides can	
CC	also be used to aid wound healing, to prevent skin aging due to sunburn,	
CC	to maintain organs before transplantation, to regenerate tissues and in	
CC	chemotaxis. The polypeptides can also be used as a food additive or	
CC	preservative to increase or decrease storage capabilities. Note: The	
CC	sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences	
Query Match	37.8%; Score 1830; DB 5; Length 1830;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1830;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	201 AGATGAACAGCTTCTCTACTTTATGGATCAAAAGTGGAATGGAATTCAGATCCCCC	260
Db	1 AGATGAACAGCTTCTCTACTTTATGGATCAAAAGTGGAATGGAATTCAGATCCCCC	60
Qy	261 TTCTCACCGGCTGTCTATTAAACCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTTAA	320
Db	61 TTCTCACCGGCTGTCTATTAAACCTACTTTGGCCAGTGTGAATGAATCTGCAGTTCTTAA	120
Qy	321 TGAGTCACAACCACCACTGAAAGCTTCTTCCCTGGCTCATTCAGCTCCCTGACACACAGA	380
Db	121 TGAGTCACAACCACCACTGAAAGCTTCTTCCCTGGCTCATTCAGCTCCCTGACACACAGA	180

```
Db 1261 AGATCAGGCTCTAGATTTTAAAGGAACTTGCATTAGTGAAGATGATGATTTCTC 1320
QY 1521 CACTGTTATAGACACACAGCAGCAAAATATCTACTAATGGTGTGATTCCTATGGAAT 1580
Db 1321 CACTGTTATAGACACACAGCAGCAAAATATCTACTAATGGTGTGATTCCTATGGAAT 1380
QY 1581 GAAGACCCAGGTGTTCTTTTGTTCCTTCCAAAGACTTTACCTCCAAAGAGATTGATTAAC 1640
Db 1381 GAAGACCCAGGTGTTCTTTTGTTCCTTCCAAAGACTTTACCTCCAAAGAGATTGATTAAC 1440
QY 1641 ACAAGAAAAGAAATAGAGGAAGCAAGTCAAGTCTCAATATTTATGAACAGAG 1700
Db 1441 AGAAGAAAAGAAATAGAGGAAGCAAGTCAAGTCTCAATATTTATGAACAGAG 1500
QY 1701 AGGAATAGAGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGTGACCTAATGAA 1760
Db 1501 AGGAATAGAGCCACAGAGGAGTGGACTACTTTTAAACAGCACTGTGACCTAATGAA 1560
QY 1761 GAAAAATATTACATAATTTCTGTAGTCAAGTTCATTCAGTGTGGGCAATCTTCCCC 1820
Db 1561 GAAAAATATTACATAATTTCTGTAGTCAAGTTCATTCAGTGTGGGCAATCTTCCCC 1620
QY 1821 CAAGGTAGTACCAAGCTTGCATCTATCAGTGTTCCTTTTGGTGTGCAAGACCCAAAGCA 1880
Db 1621 CAAGGTAGTACCAAGCTTGCATCTATCAGTGTTCCTTTTGGTGTGCAAGACCCAAAGCA 1680
QY 1881 ACCTTCTAATCTTAACTTCAAAATTCAAAGCCATTTATCAGACCAATTTACAAATGACTT 1940
Db 1681 ACCTTCTAATCTTAACTTCAAAATTCAAAGCCATTTATCAGACCAATTTACAAATGACTT 1740
QY 1941 TCCTCAACAGTGAATATATCTATAAATAAATATCTTGGGAAGCAAAAT 2000
Db 1741 TCCTCAACAGTGAATATATCTATAAATAAATATCTTGGGAAGCAAAAT 1800
QY 2001 AGGGAAAACCTCAGCAACCAATGTATGCAG 2030
Db 1801 AGGGAAAACCTCAGCAACCAATGTATGCAG 1830

RESULT 5
ADB33428
ID ADB33428 standard; DNA; 1830 BP.
XX
AC ADB33428;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human novel lung related polypeptide DNA SEQ ID NO 355.
XX
KW gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
KW adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
KW immunodeficiency; X-linked agammaglobulinaemia;
KW X-linked infantile agammaglobulinaemia; inflammatory disorder;
KW adrenaritis; alveolitis; immune complex disease; serum sickness;
KW polyarthritis nodosa; bleeding disorder; thrombocytopenia;
KW Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
KW multiple myeloma; macrophage related disorder; Gaucher's disease;
KW Neimann-Pick disease; tumour; colon cancer; pancreatic cancer;
KW renal disorder; nephritis; bone disorder; Albers-Schonberg disease;
KW bowleg; muscle disorder; Becker's muscular dystrophy;
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;
KW traumatic lesion; endocrine disorder; Cushing's syndrome;
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;
KW gastric reflux; human; ds.
XX
OS Homo sapiens.
XX
FN US2003054368-A1.
XX
PD 20-MAR-2003.
XX
XX 22-FEB-2002; 2002US-00079854.
XX
```

```
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189574P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229143P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234598P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
```

PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249298P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2000US-0254097P.
 PR 17-JAN-2001; 2001US-00259678P.
 PR 17-JAN-2001; 2001US-00754878.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-695900/66.
 XX Novel isolated lung antigen polypeptides useful for treating, preventing,
 PT

PT diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
 PT Von Willebrand's disease.
 XX Disclosure; SEQ ID NO 355; 178pp; English.
 XX The invention relates to an isolated lung antigen polypeptide sequence or
 CC encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
 CC are useful for treating, preventing, diagnosing and/or prognosing
 CC diseases and/or disorders such as pathological cell proliferative
 CC neoplasias e.g. acute myelogenous leukaemias, adenocarcinoma; respiratory
 CC disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
 CC X-linked agammaglobulinemia, X-linked infantile agammaglobulinemia;
 CC inflammatory disorders such as adenitis, alveolitis; immune complex
 CC diseases such as serum sickness, polyarthritis nodosa; bleeding disorders
 CC such as thrombocytopenia, Von Willebrand's disease; acquired platelet
 CC dysfunction such as kidney failure, multiple myeloma; disorders
 CC associated with macrophage numbers and/or macrophage function such as
 CC Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
 CC pancreatic cancer; renal disorders such as kidney failure, nephritis;
 CC bone disorders such as Albers-Schonberg disease, bowlegs; muscle
 CC disorders such as Becker's muscular dystrophy, Duchenne's muscular
 CC dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
 CC; endocrine disorders such as Cushing's syndrome, corticosteroid

Query Match 37.8%; Score 1830; DB 9; Length 1830;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 AGATCAAAACAGTTCTTCTACTTTATTGGATACAAAGTGAATGAATTTCTAGATCCCC 260
 DB 1 AGATCAAAACAGTTCTTCTACTTTATTGGATACAAAGTGAATGAATTTCTAGATCCCC 60
 QY 261 TTCTCACCGGTGTCATTTACCTTCTTGGCCAGTGTGAATGAATCTCAGTTCTAA 320
 DB 61 TTCTCACCGGTGTCATTTACCTTCTTGGCCAGTGTGAATGAATCTCAGTTCTAA 120
 QY 321 TGAGTCACACCAACCACTGAAGTCTTCTCCCTGGCTCATTCAGCTCCCTGACCACAGA 380
 DB 121 TGAGTCACACCAACCACTGAAGTCTTCTCCCTGGCTCATTCAGCTCCCTGACCACAGA 180
 QY 381 GGAGAGGATCACTGTGCTAATGACAGGACTGTAATCTAATCCAGAGATGCCACAAT 440
 DB 181 GGAGAGGATCACTGTGCTAATGACAGGACTGTAATCTAATCCAGAGATGCCACAAT 240
 QY 441 GTGGATTGATGAAATGCTGTTCGAGAGACCACTTAATTAAGAGAACTATAGTTGGGA 500
 DB 241 GTGGATTGATGAAATGCTGTTCGAGAGACCACTTAATTAAGAGAACTATAGTTGGGA 300
 QY 501 TGATCAATGCAGTCTGTTCGAGTGGGAGAGAGAAATGCGAAACCTGGCTTGTCTGCC 560
 DB 301 TGATCAATGCAGTCTGTTCGAGTGGGAGAGAGAAATGCGAAACCTGGCTTGTCTGCC 360
 QY 561 AGATGAGAAGAAATGTTCTTGTGTAGCCGTCACTGATCACTGTGATAAAAGGACATTACA 620
 DB 361 AGATGAGAAGAAATGTTCTTGTGTAGCCGTCACTGATCACTGTGATAAAAGGACATTACA 420
 QY 621 AAACGATTTACAGGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 680
 DB 421 AAACGATTTACAGGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
 QY 681 TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTATAGTATAAATCAGTCCAC 740
 DB 481 TTCACTGGATAATGAAACACAGACAACTGATCAATTTAGTTTATAGTATAAATCAGTCCAC 540
 QY 741 TGAAGAGATATGAATTCAGAGAAACAAATGGATCCATTTGAATAGCCGAAACACAGGG 800
 DB 541 TGAAGAGATATGAATTCAGAGAAACAAATGGATCCATTTGAATAGCCGAAACACAGGG 600
 QY 801 GAGATCTGTTAAACCATCTGTGTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 860
 DB 601 GAGATCTGTTAAACCATCTGTGTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 660
 QY 861 TTCACATTAAGGATCAGCGAGATAGGTAGAGACCCCTCCATGCTCTCGGATTACAAG 920

Db 661 TTACACATTTAAAGATGACGAGATATAGTATAGAGACCCCTCCATGCTCGGATTAACAG 720
QY 921 TTAAACGGTTGATTCAGTAATCTATCCAGGGAACAGATGGATGCTGCTGTTTAAAAA 980
Db 721 TTAAACGGTTGATTCAGTAATCTATCCAGGGAACAGATGGATGCTGCTGTTTAAAAA 780
QY 981 GCAAGAGACTATATACAGATGAGGACCTACTGGGAAATCAGCTCTCTAGACAGA 1040
Db 781 GCAAGAGACTATATACAGATGAGGACCTACTGGGAAATCAGCTCTCTAGACAGA 840
QY 1041 TCTAGGGAGTCCAAATTCCTTTTCCCATGATGATGAGGGGATTTTGAAGAAAAAGAGCC 1100
Db 841 TCTAGGGAGTCCAAATTCCTTTTCCCATGATGATGAGGGGATTTTGAAGAAAAAGAGCC 900
QY 1101 ACCAGAGGAGACCACTGAGAAATCCCTCGGTCTGGTTTACCTTTGCTTCTCAAAAC 1160
Db 901 ACCAGAGGAGACCACTGAGAAATCCCTCGGTCTGGTTTACCTTTGCTTCTCAAAAC 960
QY 1161 AGACATGCTTAATGGTCTGGAAGGAATAATGACTGCAACGGTGTTCAGATTGCTTGT 1220
Db 961 AGACATGCTTAATGGTCTGGAAGGAATAATGACTGCAACGGTGTTCAGATTGCTTGT 1020
QY 1221 GCCTAATGAAGTTAGGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 1280
Db 1021 GCCTAATGAAGTTAGGGCTGATGAAATGAAGTTATGAACATGAAGAACTCTTGGCAC 1080
QY 1281 TACAGAAATCCCTTAATATGACAGAGCAATTCCTGAACTCTCAGACATGACTAATGGAA 1340
Db 1081 TACAGAAATCCCTTAATATGACAGAGCAATTCCTGAACTCTCAGACATGACTAATGGAA 1140
QY 1341 GTTGACTAACTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db 1141 GTTGACTAACTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1401 ACAGATTAGTACGCTGAGGACACTAATGATGATGATGATGATGATGATGATGATGATG 1460
Db 1201 ACAGATTAGTACGCTGAGGACACTAATGATGATGATGATGATGATGATGATGATGATG 1260
QY 1461 AGATCCAGGTCTAGATTTAAAGAACTTGTGATGATGATGATGATGATGATGATGATGATG 1520
Db 1261 AGATCCAGGTCTAGATTTAAAGAACTTGTGATGATGATGATGATGATGATGATGATGATG 1320
QY 1521 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGATGATGATGATGATGATGATG 1580
Db 1321 CACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGATGATGATGATGATGATGATG 1380
QY 1581 GCAAGACCCAGGTGTTCTTTTGTTCAGAGACTTTTACCTCCAGAGCAATTCAGTAAAC 1640
Db 1381 GCAAGACCCAGGTGTTCTTTTGTTCAGAGACTTTTACCTCCAGAGCAATTCAGTAAAC 1440
QY 1641 AGAAGAAAAAATAGAGGAAACCAAGTCAAGATGCTACTCAAAATATTTATGAACAGAG 1700
Db 1441 AGAAGAAAAAATAGAGGAAACCAAGTCAAGATGCTACTCAAAATATTTATGAACAGAG 1500
QY 1701 AGGAATGAGCCACAGAGGAGTGAATCTTTTAAACAGCACTGCTGCTGCTGCTGCTGCTG 1760
Db 1501 AGGAATGAGCCACAGAGGAGTGAATCTTTTAAACAGCACTGCTGCTGCTGCTGCTGCTG 1560
QY 1761 GAAAAATATTTTACATAATTTCTGTAGTCAAGTTTCCATCAGTGTGGGCAATCTTCCCC 1820
Db 1561 GAAAAATATTTTACATAATTTCTGTAGTCAAGTTTCCATCAGTGTGGGCAATCTTCCCC 1620
QY 1821 CAAGGTAGTACAGCCTGCACTATCAGTGTCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGG 1880
Db 1621 CAAGGTAGTACAGCCTGCACTATCAGTGTCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGG 1680
QY 1881 ACCTTCTAAATCTTAAATCTCAAAATCCAAAGCAATATCAGACCAATTTACAAATCAGT 1940
Db 1681 ACCTTCTAAATCTTAAATCTCAAAATCCAAAGCAATATCAGACCAATTTACAAATCAGT 1740
QY 1941 TCTGCAAAACAGTGGGAAATATCTAAAAATAAAATGATATTTCTTTGGGAAAGCAAAAT 2000

Db 1741 TCTGCAAAACAGTGGGAAATATCTAAAAATAAAATGATATTTCTTTGGGAAAGCAAAAT 1800
QY 2001 AGGGGAAATCTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 2030
Db 1801 AGGGGAAATCTCAGCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1830
RESULT 6
ADB33429
ID ADB33429 standard; DNA; 1830 BP.
XX AC ADB33429;
XX XX
XX 04-DEC-2003 (first entry)
XX
XX Human novel lung related polypeptide DNA SEQ ID NO 356.
KW gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
KW adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
KW immunodeficiency; X-linked agammaglobulinemia;
KW X-linked infantile agammaglobulinemia; inflammatory disorder;
KW adrenitis; alveolitis; immune complex disease; serum sickness;
KW polyarteritis nodosa; bleeding disorder; thrombocytopenia;
KW Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
KW multiple myeloma; macrophage related disorder; Gaucher's disease;
KW Neimann-Pick disease; tumour; colon cancer; pancreatic cancer;
KW renal disorder; nephritis; bone disorder; Albers-Schonberg disease;
KW bowler; muscle disorder; Becker's muscular dystrophy;
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;
KW traumatic lesion; endocrine disorder; Cushing's syndrome;
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;
KW gastric reflux; human; ds.
XX
XX Homo sapiens.
OS
XX
XX US2003054368-A1.
XX
XX 20-MAR-2003.
XX
XX 22-FEB-2002; 2002US-00079854.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190075P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.


```
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 03-SEP-2000; 2000US-0229509P.
PR 03-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234233P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234938P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241783P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764878.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-695900/66.
XX
XX Novel isolated lung antigen polypeptides useful for treating, preventing,
PT diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
PT Von Willebrand's disease.
XX
XX Disclosure; SEQ ID NO 356; 178pp; English.
XX
XX The invention relates to an isolated lung antigen polypeptide sequence or
CC encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
CC are useful for treating, preventing, diagnosing and/or prognosing
CC diseases and/or disorders such as pathological cell proliferative
CC neoplasias e.g. acute myelogenous leukemias, adenocarcinoma; respiratory
CC disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
CC X-linked agammaglobulinaemia, X-linked infantile agammaglobulinaemia;
CC inflammatory disorders such as adenitis, alveolitis, immune complex
CC diseases such as serum sickness, polyarteritis nodosa; bleeding disorders
CC such as thrombocytopenia, Von Willebrand's disease; acquired platelet
CC dysfunction such as kidney failure, multiple myeloma; disorders
CC associated with macrophage numbers and/or macrophage function such as
CC Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
CC pancreatic cancer; renal disorders such as kidney failure, nephritis;
CC bone disorders such as Albers-Schönberg disease; bowlegs; muscle
CC disorders such as Becker's muscular dystrophy, Duchenne's muscular
CC dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions
CC ; endocrine disorders such as Cushing's syndrome, corticosteroid
Query Match 37.8%; Score 1830; DB 9; Length 1830;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1830; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 201 AGATGAACAGTTCTTCTACTTTATGGATACAAGTGGATGAATAGATTCTAGATCCCC 260
Dd 1 AGATGAACAGTTCTTCTACTTTATGGATACAAGTGGATGAATAGATTCTAGATCCCC 60
Qy 261 TTCTACCGGCTGTCATTTAAACCTACTTTGGCCAGTGTGAATGATCTGCAGTTTCTAA 320
```

Db 61 TTCTCACCCTGTCTCAATTAACCTTACCTTGGCCAGTGTGAATGAATCTGCAGTTCTTAA 120
QY 321 TGAGTCACACCACTCACTCAAAAGTCTTCTCCCTGCTCATTCAGCTCCCTTGACACAGA 380
Db 121 TGAGTCACACCACTCACTCAAAAGTCTTCTCCCTGCTCATTCAGCTCCCTTGACACAGA 180
QY 381 GGAAGAGGATCACTGTGCTTAATGGACAGCACTGTAACTTAAATCCAGAGATTGCCACAAT 440
Db 181 GGAAGAGGATCACTGTGCTTAATGGACAGCACTGTAACTTAAATCCAGAGATTGCCACAAT 240
QY 441 GTGATTGATGAATGCTGTTGCGAGACAGCACTTAATTAAGAGAACTATAGTTGGGA 500
Db 241 GTGATTGATGAATGCTGTTGCGAGACAGCACTTAATTAAGAGAACTATAGTTGGGA 300
QY 501 TGATCAATCCAGTGTGTTGAAAGTGGAGAGAGAAATGTGGAACCTGGCTTGTCTGCC 560
Db 301 TGATCAATCCAGTGTGTTGAAAGTGGAGAGAGAAATGTGGAACCTGGCTTGTCTGCC 360
QY 561 AGATGAGAAGATGTTCTTGTGAGCCGTCATGATCACTGTGATGAATGAATGAGTCCAC 620
Db 361 AGATGAGAAGATGTTCTTGTGAGCCGTCATGATCACTGTGATGAATGAATGAGTCCAC 420
QY 621 AAACGATTACAGGATGTGAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 680
Db 421 AAACGATTACAGGATGTGAATTAATTAATAGTCAATCCCTTATGATGCTTTTAGCTG 480
QY 681 TTCATCGGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 740
Db 481 TTCATCGGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 540
QY 741 TGAAAGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 800
Db 541 TGAAAGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 600
QY 801 GAGATCTGTGAACCTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 860
Db 601 GAGATCTGTGAACCTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 660
QY 861 TTCACAAATGAAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 920
Db 661 TTCACAAATGAAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 921 TTTAAGGTTGATTCAGTAAATCTCATCCAGGACAGATGATGATGATGATGATGATGATGAT 980
Db 721 TTTAAGGTTGATTCAGTAAATCTCATCCAGGACAGATGATGATGATGATGATGATGATGAT 780
QY 981 GCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTAGGACAGA 1040
Db 781 GCAAGAGAACTATATACAGATGAGGACCTCACTGGCAAAATCAGCTCTCTAGGACAGA 840
QY 1041 TCTAGGGATCCAAATCTCTTCCCATGAGTGGGATTTGATGAAGAGGAC 1100
Db 841 TCTAGGGATCCAAATCTCTTCCCATGAGTGGGATTTGATGAAGAGAGGAC 900
QY 1101 AGCAGAGAGACCACTGAGAAATCCCTCCGCTCTGTTTACCTTTGCTTCTCAAAAC 1160
Db 901 AGCAGAGAGACCACTGAGAAATCCCTCCGCTCTGTTTACCTTTGCTTCTCAAAAC 960
QY 1161 AGACATGCTTAATGGTCTGAAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 1220
Db 961 AGACATGCTTAATGGTCTGAAGAAATATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1221 GCCTAATGAGTTAGGCTGATGAATGAAGGTTATGAACATGAAGAACTCTTGCCAC 1280
Db 1021 GCCTAATGAGTTAGGCTGATGAATGAAGGTTATGAACATGAAGAACTCTTGCCAC 1080
QY 1281 TACAGAACTCTTAATGACAGAGCAATTTCTCTGAATCTCAGACATGATGATGATGATGATGAT 1340
Db 1081 TACAGAACTCTTAATGACAGAGCAATTTCTCTGAATCTCAGACATGATGATGATGATGATGAT 1140
QY 1341 GTTGACTAAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400

Db 1141 GTTGACTAAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1401 ACAGATTAGTCAGCTGAGGACACTAATGGTGTAGTGTGAGGACAGTGTGTGGATTGGC 1460
Db 1201 ACAGATTAGTCAGCTGAGGACACTAATGGTGTAGTGTGAGGACAGTGTGTGGATTGGC 1260
QY 1461 AGATGCAAGTCTAGATTAAAAGGAACTTGCATTAGTGAAGTGAAGAAATGTGATTCTC 1520
Db 1261 AGATGCAAGTCTAGATTAAAAGGAACTTGCATTAGTGAAGTGAAGAAATGTGATTCTC 1320
QY 1521 CACTGTATAGACACACAGCAGCAATTAATCTATCTAATGTTGTGATCTCTATGGAAT 1580
Db 1321 CACTGTATAGACACACAGCAGCAATTAATCTATCTAATGTTGTGATCTCTATGGAAT 1380
QY 1581 GCAAGACCCAGGTTCTTTTCTTCCAAAGACTTTTACCTCCAAAGAGATTTCAGTAAC 1640
Db 1381 GCAAGACCCAGGTTCTTTTCTTCCAAAGACTTTTACCTCCAAAGAGATTTCAGTAAC 1440
QY 1641 AGAAGAAAAGAAATAGAGAAAGCAAGTCAGAATGCTACTCAAAATATTTATGAACAGAG 1700
Db 1441 AGAAGAAAAGAAATAGAGAAAGCAAGTCAGAATGCTACTCAAAATATTTATGAACAGAG 1500
QY 1701 AGGAATGAGGCCACAGAGAGGAGTGACTCTTTTAAACAGCACTGGTGACCTAATGAA 1760
Db 1501 AGGAATGAGGCCACAGAGAGGAGTGACTCTTTTAAACAGCACTGGTGACCTAATGAA 1560
QY 1761 GAAAAATTTATACATAATTTCTGTAGTCAAGTTCATCAGTGTGGGCAATCTTCCCC 1820
Db 1561 GAAAAATTTATACATAATTTCTGTAGTCAAGTTCATCAGTGTGGGCAATCTTCCCC 1620
QY 1821 CAAGTAGTAGCAAGCTGCCATCTATCAGTGTCTTGTGTGTGCAAGCCCAAGCA 1880
Db 1621 CAAGTAGTAGCAAGCTGCCATCTATCAGTGTCTTGTGTGTGCAAGCCCAAGCA 1680
QY 1881 ACCTTCTAATCTTAAACTTCAAACTTCAAAAGCCATTATCAGACCATTTACAAATGACTT 1940
Db 1681 ACCTTCTAATCTTAAACTTCAAACTTCAAAAGCCATTATCAGACCATTTACAAATGACTT 1740
QY 1941 TCTGCAAAACAGTGGAAATATATCTAATAAATAAATAATGATATCTTGGAAAGCAAAAT 2000
Db 1741 TCTGCAAAACAGTGGAAATATATCTAATAAATAAATAATGATATCTTGGAAAGCAAAAT 1800
QY 2001 AGGGGAAAACCTCAGCAACCAATGTATGAG 2030
Db 1801 AGGGGAAAACCTCAGCAACCAATGTATGAG 1830

RESULT 7
AAZ50070

ID AAZ50070 standard; cDNA; 4823 BP.

XX

AC AAZ50070;

XX

DT 04-MAY-2000 (first entry)

XX

Xenopus Smad Anchor for Receptor Activation protein-1 encoding cDNA.

XX

Smad Anchor for Receptor Activation protein; XSAR1; TGF-beta;
transforming growth factor-beta; bone morphogenetic protein; BMP;
activin; anti-inflammatory; cytostatic; anarthritic; vulnarary;
TGF-beta modulator; wound healing; scarring; arthritis; immune response;
inflammatory response; tumour progression; cell proliferation; fibrosis;
Smad protein; fibrogenesis; tissue morphogenesis; tissue damage; ss.

OS Xenopus sp.

XX

Key Location/Qualifiers
316..4023
/*tag= a

FH CDS

FT /product= "Xenopus SARA-1 protein"
FT /note= "Binds to receptor regulated Smad proteins"

XX

PN WO200005360-A1.

XX 03-FEB-2000.
 PD 20-JUL-1999; 99WO-CA000656.
 PF 20-JUL-1998; 98CA-02237701.
 PR 10-DEC-1998; 98CA-02253647.
 XX (HSCR-) HSC RES & DEV LP.
 PA Wrana JL;
 XX WPI; 2000-182691/16.
 DR P-PSDB; AAY44751.
 DR
 XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 PT developing agents for treating e.g. wounds, arthritis, fibrosis,
 PT inflammation, tumors, fibrogenesis or tissue damage.
 XX
 PS Claim 9; Page 57-58; 93pp; English.
 XX
 CC The present sequence is the cDNA encoding the Xenopus Smad Anchor for
 CC Receptor Activation protein-1 (XSARA), a TGF-beta modulator. It is
 CC isolated from lambda ZAP II xenopus dorsal lip library and has 62%
 CC identity to hSARA1. The XSARA proteins bind to receptor-regulated Smad
 CC proteins and ensures its appropriate localisation for activation by a
 CC type I receptor of a TGF-beta, activin or bone morphogenetic protein
 CC (BMP) signalling pathway. SARA proteins has anti-inflammatory,
 CC cytostatic, anti-arthritis and vulnary activity. The SARA proteins are
 CC useful for developing agents for prevention or treatment of disorders
 CC involving TGF-beta superfamily member signalling pathway. Such agents can
 CC be used for wound healing, scarring, arthritis and fibrosis (e.g. liver
 CC and kidney), in modulating inflammatory and immune responses, tumour
 CC progression, cell proliferation or fibrogenesis and in tissue
 CC morphogenesis. It is also used for protecting, restoring and regenerating
 CC tissues after tissue damage
 XX
 SQ Sequence 4823 BP; 1405 A; 1035 C; 1089 G; 1294 T; 0 U; 0 Other;
 Query Match 31.9%; Score 1545.8; DB 3; Length 4823;
 Best Local Similarity 71.8%; Pred. No. 0;
 Matches 2136; Conservative 0; Mismatches 792; Indels 45; Gaps 7;
 QY 1443 ACAGTGTGTGATGGGAGATGAGGCTAGATTTAAAGCACTTCGATTAGTGAAG 1502
 DB 1089 AGAGTTTGAGAGGGGTAGATGCTTAATGCTTTGTTGGAATGATGATGTTATAA 1148
 QY 1503 TGAAGAATGTGATTTCTCCACTGTTATAGACACACAGCAGCAAAATATCTATCTAATGG 1562
 DB 1149 AGAAGCAATAGATTTGGCTGAAGAAATGGAACCTAATGCACCAATGCTCTGTACAAATGG 1208
 QY 1563 TTGTGATTCCTATGGAATGCAAGCCAGGCTTTCTTTTCTCCAAAGCTTTACCTC 1622
 DB 1209 GTGTGATTCCTATGGAATGAAACCCAGCGTAGCTCAAAACCAAGAAATTTACCTTC 1268
 QY 1623 CAAGAAGATTCAGTAACAGAGAAAGAAAGAAATAGAGAAAGCAAGTCAAGTCTACTC 1682
 DB 1269 AAAAGAAGATTCGTGACAGAGAAAGAAAGAAATGGAAGAAAGTCAAGTCAATATAC 1328
 QY 1683 ABATATTTATG---RACAGAGGGAATGAGGACACAGAGGAGTGGACTCTTTTAA 1739
 DB 1329 TGGTGTGTTATGAACACCAAGAGAGATGATGTTACAGAGAGAGTGGACTCTCTGTATA 1388
 QY 1740 CAGCACTGGTGAATTAATGAAGAAATATTTATACATAATTTCTGTAGTCAAGTTCATC 1799
 DB 1389 TGCTAAGGCTGACCAATGAGAACCA---ATTGCTAGTCTTTGTAATCAGGTTCCATC 1445
 QY 1800 AGTGTCTGGGCAATCTTCCCAAGTAGTAGCAGGCTGCACTATCATGATGTTCTTT 1859
 DB 1446 CATGATGGGCAACATCACCACCAAGAGGCAAGTGTGCAATCTCTCAGTGTTCATA 1505
 QY 1860 TGGTGGTGCAGACCCCAAGCAACCTTCTAATCTTAAACTTCAATTCCAAAGCCATTATC 1919

DB 1506 CGGTGAGCAGCAGCACTAAGCAGCACTCATCTCAAATCCCATATTTCCAAAGCCATTGTC 1565
 QY 1920 AGACCATTTACAAA---ATGACTTTCTCTGCAACAGTGGAATAATACTATAAAATAAAAA 1976
 DB 1566 TGAATGTTGACAGCGATCTCAATTCCTCAAATGCTGGTGCAGCTCTAATATCAAAAA 1625
 QY 1977 TGATATTTCTGGGAAGCAAAATTAGGGGAAATCTAGCAACCAATATGATGATGATGATC 2036
 DB 1626 TGACATGTTAAACAAATCAAAATCAGGGGGATAAAGCTGATTTTCAAGATCACTCGCTG 1685
 QY 2037 TTTGGAAACATCTCTAATGTCGATACAAATAGGGGCAACATTTAGAAAGTTATAGAGCTGA 2096
 DB 1686 TTCTGACGTGCGCAGCCCTGT---TACTGATGCTAATGGTGATTTCCCTGGAGATACAG 1742
 QY 2097 GATCTCCACTAGACCATGCTTGTGATTTAGCTCCAGATAGCCAGATATGATTTCTCAGAGC 2156
 DB 1743 GGGACCTGCGACCTTTGCTTGTGAGTGTCTCCAGACAGCCAGCAACCAATCTGCTTGC 1802
 QY 2157 TGGTCAGTTTGGAAATTTCTGCCAGAAAGCCATTACCAACGCTGGGTGAGGGTCTCCAGT 2216
 DB 1803 CGGGCAGTTGGGTACCCATCTCTAAGCCATTTACTACTTAGGGAGTGGCTCCAGT 1862
 QY 2217 ATGGGTACGGGATTTCTAGGCTCCAAATGTGATGAATGTGAAGCCAGGTTTACATTCAC 2276
 DB 1863 CTGGGTGCGAGATTTCCCAAGCACCACCAACTGATGAAGTGCAGGCCAGATTTTACATTTAC 1922
 QY 2277 CAAAAGGAGGCATCACTGCAGAGCATGTGGGAAGGTTTCTGTCTCTCTCTCTCTCTCT 2336
 DB 1923 CAAAAGGAGGCATCACTGCCAGCTTTGGGAAGAGTGTCTGTCTCTCTCTCTCTCTCT 1982
 QY 2337 GAAATGTAACCTGTTATACATGTGACAGAAAGAGCTAGAGTGTGTATCTTCCCATTC 2396
 DB 1983 AAAATGCCAACTACAGTACATGATGATAAAAGAGGCTCGTGTGTGTGTATTTCTCATTC 2042
 QY 2397 AGTGTCTAATGATGCTCAAGCTTGGGAGAACATCATGATGCTCCCAAGCCAGAGCCCTAA 2456
 DB 2043 TGTGCTTAATGATGCTCAAGCATGGGAGAACATTTAAGTCATCGTCTCAAGCCCAA 2102
 QY 2457 CCTAACCAATCTCTGCTGAATCTGTTCTACTATCCCTCCCTTTCAGCAAGCTCAGGCTC 2516
 DB 2103 TCCAAATATCTGCTGATTAATCTGCTCAACTATCCCTCCGATGAGCAGGCAAGCTTC 2162
 QY 2517 AGGACCTGTAGCTCTCCACCTCCACCTGTGATGATGATGATGATGATGATGATGATGAT 2576
 DB 2163 AGGACCTGTAGTTCCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCT 2222
 QY 2577 TGGACAGAAAGTGGCTCAGCCAGAGCAGAGCAGAGCGAGTTTGGTTTCTGATGGATCTT 2636
 DB 2223 AGGAACTGAAGGTCACAGTCAAGGAAACAGCGCGCTGTTTGGTTTCTGATGGAATAT 2282
 QY 2637 GCCCAATGAGAGTGTGCTGATGAGCCAAATTAACATGAATGGAATCTCTCTGCGAG 2696
 DB 2283 ACCCAAGGAGAGTGTGCTGATGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2336
 QY 2697 AACCTGTGTGTGTCACAGCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 2756
 DB 2337 GACACTTACTGTGTCACATACC-----AACAAATCCACATCTTTCAGAGTC 2381
 QY 2757 GACGATATTTGCTCTATTTCTCTGGAGTATACTCAGTTTGGAGTCTCTCTGAGAGTGC 2816
 DB 2382 TGAGAACCTCTGGAATTTCTGTGGAAGTATACTCAG-----GTTGCGAGTGC 2429
 QY 2817 AATGAATCTTATCTCTGAAGTGGCTTCTCTCCCATTTCTCATCTCCACTCGTGTAAAGG 2876
 DB 2430 AATGAACCTTATCCAGAAAGATGGCTTCTCTCTTACTAATCTCTACTGGAGTAAAGG 2489
 QY 2877 AGACTATCTGTGAAGAGAAACCATCAGATTTCAATTAATCAGCAGTGGAGGATGG 2936
 DB 2490 AGATTACGAGTTGAGGAACGCCCTTCCAGATGCTGTGATGAGCAACTAGAGGAAGG 2549
 QY 2937 TGGCCCTCAGCCACTGTGTATTTGTTTAAATGCAAAATTTGTGTCATTTGTTTAAATTTGT 2996
 DB 2550 AGGACCAATCCTTGGTTTTGTTTCTTAATGCAAAATCTTTTGGCCATGTTTGAATCGT 2609

QY 2997 AAATTATGTGAACAGGAAGTGTGGTGTTCACAAAGGGAATGCATGAGTGGGTCA 3056
Db 2610 GAACTATGTTAAACAGGAATGCTGGTGTCTTACTACAAAGGAATGCATGAGTGGGCA 2669
QY 3057 GTCTGAGATAGTCAATCTTCTACAGTGTTCACGAGTAAAGTGTTCGCCAAAGATAT 3116
Db 2670 GGTGAGATCGTAATCTCTTTCAGTGTCTGCTGATGAGAGTGTCTGCGGAGGACCT 2729
QY 3117 CTTTAATCACTTGTGCACCTTATCGGATCTCTGGCAGGGAATGTGTGAGCAACTT 3176
Db 2730 GTTTAGCCATTTTGTGAGCTGTATCAGAGGCAATTCAGAGTAATGTAGTGGGAACCT 2789
QY 3177 GGGACATTCCTTCTTCAGTCAAAAGTTTCTTGGCAGTAAAGAACATGTGTGATTTTATA 3236
Db 2790 GGGGCATTCCTTCTCAGCCAGAGTTTCTGGGTAGTAAGGATCATGGTGGATTTCTTTA 2849
QY 3237 TGTGACATCTACCTACAGTCACTGCAGACCTAGTACTCCCAACCCCACTTACTTGT 3296
Db 2850 TGTTCGACCAACTTACAGTCCCTCCAGACCTGTCTTCTTCTGAGAGCCGTACTTGT 2909
QY 3297 TGGGATTCCTTATCCAGAAATGGAAACTCTTGGCTAAAGTATTTCTTATCCGTCTGAT 3356
Db 2910 TGGATTCCTTATCCAAAGTGGAGACTCCATGGGCCAAAGTTCCTCCCATTCGGCTTAT 2969
QY 3357 GTTGAGACTTGGAGCTGAATATCGACTTTATCCATGCCCACTATTCAGTGTGATTCG 3416
Db 2970 GCTGCGTTAGTGTGAGAAATACAGATTTGACCCATGTCCACTCTTCACTGTTCGATACAG 3029
QY 3417 GAAGCCATTTTGGAGAGAGCGGGCATACCATCATGAATCTTCTTGAGACTTTCAGAAA 3476
Db 3030 AAAACCTCTGTTTGGGAAACCGGACACACCATCATTAATGTTTACGCCGATTTTCAGAAA 3089
QY 3477 TTACAGTATACCTTGGCAGTAGTTCAAAGTTTGGTGTTCATATGGAAGTTTCGAAAC 3536
Db 3090 CTATCAGTATATCTTCCAGTGTGTGGAGGCTTGGTGTGATATGGAAGTTCAGAAAAC 3149
QY 3537 TAGCATCAAAATTCCTCAGCAACAGATACAATGAGATGATGAAGCCATGACAGTCCAA 3596
Db 3150 TAGCAATTAATTCCTCCAGCAATAGATACAATGAGATGATGAAGCAATGAACAAATCCAA 3209
QY 3597 TGAGCATCTCTGGCAGAGTGCTCTTCAATGAAAGGAGCAGACTCTCATCTTGTGTG 3656
Db 3210 TGAGCATGTGTGGCCATAGGAGCATGCTTCAACAGATGGCAGACTCTCACTTGTGTG 3269
QY 3657 TGTACAGATGATGATGAAACTATCAGCCAGGCTATCAGTATTCACATCAGCCAG 3716
Db 3270 TGTGCAAAACGATGATGGCAATATCCAGACCCAGGCAATAGTATCCAAACCAACCCAG 3329
QY 3717 AAAAGTCACTGGTCCAGTCTTCTTGTGTTCAGTGGGCTCTGAAATCTCTCTTGATA 3776
Db 3330 TAAAGTCAACGGGCCAGTCTTCTTGTCTTCACTGAGTGGTCACTAAAGTCTCTCTCGGATA 3389
QY 3777 CTTGCCAAGTCCAGTATTTGGAAGTATGTTTATGTCAGATTCAGTATTCAGCAAGACAT 3836
Db 3390 CTTGCCAAGTCCAGTATTTGGAAGTATGTTTATGTCAGATTCAGTATTCAGCAAGACAT 3449
QY 3837 GCATTCCTTGGAGGAGGAGTGGCAGAGATGAAGGACTTCACCATCACTCTGGGAGGC 3896
Db 3450 GGATGCCCTCAGACAGTCCCTTCCGGAGATGAAGGATTCACCATTCATGTCGGAAGC 3509
QY 3897 GGAGCGGAGGAACCCGAGGAGACATCCACATCCAGTGGGTGATGATGACAGAACGT 3956
Db 3510 TGATGACAGAGGAGTACAGGAACATGCTTATGTCAGTGGGTGAGGATGACAGAACCT 3569
QY 3957 TAGCAAGGTGTCTTAACTAGTCTTATGAGTGGAGTCCATGAGACTATACAAATGTGA 4016
Db 3570 TAACAAAGGAGTCTTTTATGTCATTCGATGGCAATCAATGAGTCTGTGACCAAGCTCAA 3629
QY 4017 GATATTCATGGATCAGAAATATAAGCAATGAAAGTAAATCAGATGGACAGAGTGT 4076
Db 3630 GATTTTTCAGGCTCAGAAATACAAAGCTAGTGGAAAAATATTCGCTGATAGAGTCT 3689

QY 4077 TTTCTAGAAAACGATGACAGCACAATTTGCCTCAGTGTCTGAGATCCTGAGATCAGATGAT 4136
Db 3690 CTTTCTGGCAATGAGGAGCAACAGAGTGGCTGAGTACCCTGCTGATTCACAGCCGACT 3749
QY 4137 GACTGAGCATGTTGGCCAAAGCTTTTGGCTTGTCTCTCTCTCCTCAGCTGAAACTTCTGAA 4196
Db 3750 CACTGAAATGTGGCCAAAGCATTTCTGTTAGGCTTTGCCACACCTCAGCTACTGAA 3809
QY 4197 GGAAGATGAATGACCAACTGGAGTACGTGTGACACTTGTGAGTACAGTACAGTTGGCTA 4256
Db 3810 GGAAGATGAATGACCAAGTGTGCTGGGTGTCTGAGTACAGTACAGTGGATGGATA 3869
QY 4257 TCAAGCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGAATGATCTGATAGCCG 4316
Db 3870 CCAAGCTGGAGCAATGGGCAACTCTCTGCTGCCGATACACCAATGATTTGGATGGTGC 3929
QY 4317 CTTGGTCCGCTGATCCATGGAGGGCTGCGCAGCTTGTGAGGGCCCGTGTGATGGA 4376
Db 3930 TTTGGTACCAGTGATACACGGGGCCACATGCCAGTTAAGTGAAGGCCCTGTGATGGA 3989
QY 4377 ACTCATCTTTTATATTTCTGAAAAACATCGTATA 4409
Db 3990 GCTGATATTTTATATCTTTGAGAACATCTCTTA 4022

RESULT 8

ACA56987
ID ACA56987 standard; cdna; 888 BP.

XX ACA56987;

XX AC ACA56987;

DT 10-JUN-2003 (first entry)

XX Human adipocyte Selected Interacting domain, SID, cdna #74.

XX Human; ss; gene; prey; adipocyte; SID; selected interacting domain;
KW anorectic; antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.

XX Homo sapiens.

XX W0200286122-A2.

XX 31-OCT-2002.

XX 14-MAR-2002; 2002WO-EP003768.

XX 14-MAR-2001; 2001US-0275734P.

XX (HYBR-) HYBRIGENICS.

XX Legrain P, David L;

XX WPI; 2003-103412/09.

XX P-PSDB; ABU70443.

XX New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes.

XX Claim 7; Page 138-139; 382pp; English.

XX The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid

CC record comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are useful for
 CC preventing or treating metabolic disorders such as obesity or diabetes.
 CC The polynucleotides are useful as probes or primers. The complex is
 CC particularly useful for identifying selected interacting domains (SID
 CC (RTM)) for screening drugs that modulate the protein interaction, thus
 CC exhibiting the therapeutic effect. The present sequence encodes a SID
 CC (prey) protein of the invention

XX Sequence 576 BP; 149 A; 134 C; 138 G; 155 T; 0 U; 0 Other;

Query Match 11.9%; Score 576; DB 7; Length 576;

Best Local Similarity 100.0%; Pred. No. 6.8e-148;

Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2457 CCTAACAAATCTGCTGAATCTGTTCTACTATCCCTCCCTTGCAGCAAGCTCAGGCCTC 2516

DB 1 CCTAACAAATCTGCTGAATCTGTTCTACTATCCCTCCCTTGCAGCAAGCTCAGGCCTC 60

QY 2517 AGAGCTCTGAGCTCTCCACCTCCACCTGATGCTGAGTGGAGTTTAAACACCC 2576

DB 61 AGAGCTCTGAGCTCTCCACCTCCACCTGATGCTGAGTGGAGTTTAAACACCC 120

QY 2577 TGGAGCAGAAATGGCTCTCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2636

DB 121 TGGAGCAGAAATGGCTCTCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

QY 2637 GCCCAATGAGAGTGGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2696

DB 181 GCCCAATGAGAGTGGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

QY 2697 AACCTCTGGCTGTCTCACAGCAGCCAGTCAAGCCAGTCACTACCACTCTCTACCA 2756

DB 241 AACCTCTGGCTGTCTCACAGCAGCCAGTCAAGCCAGTCACTACCACTCTCTACCA 300

QY 2757 GACGATATTTCTTATCTCTGGAGATTAATCACTAGTGGAGTCTCTTGAAGTGC 2816

DB 301 GACGATATTTCTTATCTCTGGAGATTAATCACTAGTGGAGTCTCTTGAAGTGC 360

QY 2817 AATGAATCTTATCTGAGAGATGGCTTCTCCATCTCTCCATCTCTCCATCTCTCC 2876

DB 361 AATGAATCTTATCTGAGAGATGGCTTCTCCATCTCTCCATCTCTCCATCTCTCC 420

QY 2877 AGACTATCTCTGGAAGAGAAACCATCAAGATTTTCAATGAGAGTGGAGAGTGG 2936

DB 421 AGACTATCTCTGGAAGAGAAACCATCAAGATTTTCAATGAGAGTGGAGAGTGG 480

QY 2937 TGGCCCTGACCCCTGATTTGTTTAAATGCAAAATTTGTTCAATGTTAAATTTGT 2996

DB 481 TGGCCCTGACCCCTGATTTGTTTAAATGCAAAATTTGTTCAATGTTAAATTTGT 540

QY 2997 AAATATGTAACAGGAGTCTGTTTTCACAAC 3032

DB 541 AAATATGTAACAGGAGTCTGTTTTCACAAC 576

RESULT 10

AAS29903

ID AAS29903 standard; cDNA; 610 BP.

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

XX AC AAS29903;

KW

KW

KW

KW

XX

XX

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

fungus infection; viral infection; ocular disorder; endocrine disorder;
 gastrointestinal disorder; renal disorder; respiratory disorder;
 wound healing; skin aging; organ transplantation; food preservative;
 tissue regeneration; anti-infertility; food additive.

Homo sapiens.

WO200155303-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001301.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-0225213P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225267P.

14-AUG-2000; 2000US-0225268P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-0226279P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226868P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

KW

KW

KW

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human lung antigen cDNA polynucleotide #65.

Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;

chicken; sheep; immunosuppressive; antihypertensive; vasotropic;

antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;

cerebroprotective; neurotropic; antibacterial; virucide; fungicide; cancer;

ophthalmological; vulnerable; gene therapy; autoimmune disease; neoplasm;

hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;

cerebrovascular disorder; nervous system disorder; bacterial infection;

PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234224P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 28-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250319P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251858P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-457723/49.
 DR P-PSDB; AAU18616.
 XX Isolated polypeptide for treating, preventing and/or prognosing
 PT respiratory disorders related to the lung including lung cancers and also
 PT for testing and detection e.g. diagnosis.
 XX
 PS Claim 1; SEQ ID NO 75; 507pp; English.
 CC Sequences AAS29839-AAS29930 represent cDNA molecules, which encode the
 CC lung antigen polypeptides of the invention. Lung antigen polypeptides and
 CC their associated polynucleotides are useful in the diagnosis, treatment
 CC and prevention of various types of disorders in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological
 CC condition can be determined by detecting the presence or absence of a
 CC mutation in a lung antigen polynucleotide. The treatable disorders
 CC include autoimmune diseases such as rheumatoid arthritis.
 CC hyperproliferative disorders such as neoplasms of the breast or liver,
 CC cardiovascular disorders such as cardiac arrest, cerebrovascular
 CC disorders such as cerebral ischaemia, nervous system disorders such as
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
 CC ocular disorders such as corneal infection, endocrine disorders such as
 CC premature labour and infertility, gastrointestinal disorders such as
 CC Crohn's disease, renal disorders such as glomerulonephritis and
 CC respiratory disorders such as asthma and pleurisy. The polypeptides can
 CC also be used to aid wound healing, to prevent skin aging due to sunburn,
 CC to maintain organs before transplantation, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

Query Match 11.8%; Score 570.4; DB 5; Length 610;

Best Local Similarity 99.2%; Pred. No. 2.5e-146;

Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

QY 1 GCNACTGATCAGCAGGACTGGCTGGTGCAGCAGACATCATGAGTAAGCAGCAGA 60
 DB 1 GCNACTGATCAGCAGGACTGGCTGGTGCAGCAGACATCATGAGTAAGCAGCAGA 60
 QY 61 AGTCTCTTCTTATCATCGTGTGTAAGGGGAAAAGGTTTAAACAAGTCTCTTAAAGTGGTG 120
 DB 61 AGTCTCTTCTTATCATCGTGTGTAAGGGGAAAAGGTTTAAACAAGTCTCTTAAAGTGGTG 120
 QY 121 TTTCTCCACCGATGGAGAAATTACTTCAAGCAGAGAGCTTACACCTGGGACAGGTTTA 180
 DB 121 TTTCTCCACCGATGGAGAAATTACTTCAAGCAGAGAGCTTACACCTGGGACAGGTTTA 180
 QY 181 GATGAATTTGAACAAAACGAAGATGAAACAGATTTCTTCTACTTTATTGGATACAAAGTGG 240
 DB 181 GATGAATTTGAACAAAACGAAGATGAAACAGATTTCTTCTACTTTATTGGATACAAAGTGG 240
 QY 241 AATAAGATTTAGATCCCTTCTCCACCGGTGTGTAATTAACCTACTTTGGCAGTGG 300
 DB 241 AATAAGATTTAGATCCCTTCTCCACCGGTGTGTAATTAACCTACTTTGGCAGTGG 300
 QY 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGCTCAT 360
 DB 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACTGAAAGTCTTCTCCCTGGCTCAT 360
 QY 361 TCAGCTCCCTGACACAGAGGAGGATCACTGTGCTAATGAGCAGAGCTGTAAATCTA 420
 DB 361 TCAGCTCCCTGACACAGAGGAGGATCACTGTGCTAATGAGCAGAGCTGTAAATCTA 420
 QY 421 AATCCAGAGATTCGCCAATGATGGATTGATGAAATGCTGTTGCCAGAGCAGTTAATT 480

Db 421 AATCCAGAGATGCCACAAATGGGATTCGATGAAATGCTGTCAGAAAGACCAGTTAAT 480
QY 481 AAGAGAACTAGTGGATGATCAATCCAGTCTGTTGAAGTGGGAGAGAAATGT 540
Db 481 AAGAGAACTAGTGGATGATCAATCCAGTCTGTTGAAGTGGGAGAGAAATGT 540
QY 541 GGAACCTGGCTGTCGTCAGATCAGAGAAATGCTGTTGAGCCGTCATGCATA 598
Db 541 GG-AACTGGCTGTCGTCAGATCAGAGAAAG-K-TCTGRTGAGCCGTCATGCATA 596

RESULT 11
ADB33148
ID ADB33148 standard; cDNA; 610 BP.
XX AC ADB33148;
XX DT 04-DEC-2003 (first entry)
XX DE Human novel lung related polypeptide cDNA SEQ ID NO 75.
KW gene therapy; lung antigen; neoplasia; acute myelogenous leukaemia;
KW adenocarcinoma; respiratory disorder; chronic rhinitis; sinusitis;
KW immunodeficiency; X-linked agammaglobulinaemia; inflammatory disorder;
KW X-linked infantile agammaglobulinaemia; immune complex disease; serum sickness;
KW adrenalitis; alveolitis; bleeding disorder; thrombocytopenia;
KW Von Willebrand's disease; acquired platelet dysfunction; kidney failure;
KW multiple myeloma; macrophage related disorder; Gaucher's disease;
KW Niemann-Pick disease; tumour; colon cancer; pancreatic cancer;
KW renal disorder; nephritis; bone disorder; Albers-Schonberg disease;
KW bowleg; muscle disorder; Becker's muscular dystrophy;
KW Duchenne's muscular dystrophy; nervous disorder; ischaemic lesion;
KW traumatic lesion; endocrine disorder; Cushing's syndrome;
KW corticosteroid deficiency; gastrointestinal disorder; dysphagia;
KW gastric reflux; human; ss; gene.
XX OS Homo sapiens.
XX US 2003054368-A1.
XX PD 20-MAR-2003.
XX PF 22-FEB-2002; 2002US-00079854.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0196123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226881P.
PR 22-AUG-2000; 2000US-0226888P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 05-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232197P.
PR 14-SEP-2000; 2000US-0232198P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241828P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249364P.
PR 17-NOV-2000; 2000US-0249365P.
PR 17-NOV-2000; 2000US-0249397P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251988P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764878.
(PHUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Barash SC;
WPI; 2003-695900/66.
P-PSDB; ADB33240.

Novel isolated lung antigen polypeptides useful for treating, preventing,
diagnosing acute myelogenous leukemias, adenocarcinoma, thrombocytopenia,
Von Willebrand's disease.

Claim 3; SEQ ID NO 75; 178pp; English.

The invention relates to an isolated lung antigen polypeptide sequence or
encoded sequence in a cDNA clone. The polypeptide and its polynucleotide
are useful for treating, preventing, diagnosing and/or prognosing
diseases and/or disorders such as pathological cell proliferative
neoplasias e.g. acute myelogenous leukemias, adenocarcinoma; respiratory
disorders such as chronic rhinitis, sinusitis; immunodeficiencies such as
X-linked agammaglobulinemia, X-linked infantile agammaglobulinemia;
inflammatory disorders such as adenitis, alveolitis; immune complex
diseases such as serum sickness, polyarteritis nodosa; bleeding disorders
such as thrombocytopenia, Von Willebrand's disease; acquired platelet
dysfunction such as kidney failure, multiple myeloma; disorders
associated with macrophage numbers and/or macrophage function such as
Gaucher's disease, Niemann-Pick disease; tumours such as colon cancer,
pancreatic cancer; renal disorders such as kidney failure, nephritis;
bone disorders such as Albers-Schönberg disease, bowlegs; muscle
disorders such as Becker's muscular dystrophy, Duchenne's muscular
dystrophy; nervous disorders such as ischaemic lesions, traumatic lesions

Query Match 11.8%; Score 570.4; DB 9; Length 610;
Best Local Similarity 99.2%; Pred. No. 2.5e-146;
Matches 593; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

QY 1 GCATAGTAATCAGCAGGACTGGCTGGTGGTGACAGACATCATGATAGTAAACCCGAGA 60
DB 1 GCATAGTAATCAGCAGGACTGGCTGGTGGTGACAGACATCATGATAGTAAACCCGAGA 60
QY 61 AGTCTGTTCTTATCACGTGTGTAAAGGGGAAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
DB 61 AGTCTGTTCTTATCACGTGTGTAAAGGGGAAAAAGGTTTAAACAAGTCTCTTAAGTGGTG 120
QY 121 TTTCTCACCAGTGGAGAAATTAATTCACAGCAGAGAGCTTCAACCTGGGACAAAGTGTGA 180
DB 121 TTTCTCACCAGTGGAGAAATTAATTCACAGCAGAGAGCTTCAACCTGGGACAAAGTGTGA 180
QY 181 GATGAATTTGAACAAAACGAAGATGAACAGTCTTCTTCTACTTTATTGGATACAAAGTGG 240
DB 181 GATGAATTTGAACAAAACGAAGATGAACAGTCTTCTTCTACTTTATTGGATACAAAGTGG 240
QY 241 AATAAGATTTAGATCCCCCTTCTCACCGGTGTCAATTAACCTTACTTTGGCCAGTGTG 300
DB 241 AATAAGATTTAGATCCCCCTTCTCACCGGTGTCAATTAACCTTACTTTGGCCAGTGTG 300
QY 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACCTGAAAGTCTTCTCCCTGGCTCAT 360
DB 301 AATGAATCTGCAGTTTCTAATGAGTCACAAACCAACCTGAAAGTCTTCTCCCTGGCTCAT 360
QY 361 TCAGTCCCTGACACAGAGAGAGGATCACTGTGCTAATGGACAGACTGTAACTA 420
DB 361 TCAGTCCCTGACACAGAGAGAGGATCACTGTGCTAATGGACAGACTGTAACTA 420
QY 421 AATCCAGAGATTGCCCAATGTGGATTGATGAAATGCTTTGTCAGAGAGACAGTAAAT 480
DB 421 AATCCAGAGATTGCCCAATGTGGATTGATGAAATGCTTTGTCAGAGAGACAGTAAAT 480
QY 481 AAGAGAACTATAGTTGGGATGATCAATCCAGTGTGTTGAAGTGGGAGAGAAATGT 540
DB 481 AAGAGAACTATAGTTGGGATGATCAATCCAGTGTGTTGAAGTGGGAGAGAAATGT 540
QY 541 GGAAACCTGGCTGTCTGCCAGATGAGAGATGTTCTTGTAGCCGTGATGATA 598
DB 541 GG-AAACCTGGCTGTCTGCCAGATGAGAGAAAG-TCTGTGTGATGCCCTCATGATA 596

RESULT 12
AAZ50069 standard; cDNA; 6632 BP.
XX
AC AAZ50069;
DT 04-MAY-2000 (first entry)
XX
DE Human Smad Anchor for Receptor Activation protein-2 encoding cDNA.
XX
KW Smad Anchor for Receptor Activation protein; hSARA2; human;
KW transforming growth factor-beta; TGF-beta; bone morphogenetic protein;
KW BMP; activin; anti-inflammatory; cytostatic; antiarthritic; vulnerary;
KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;
KW inflammatory response; tumour progression; cell proliferation; fibrosis;
KW fibrogenesis; tissue morphogenesis; tissue damage; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 249..4868
FT /tag= a
FT /product= "Human SARA-2 protein"
FT /note= "Binds to receptor regulated Smad proteins"
FT /transl_except= (pos:336..338, aa:Gln)
FT /transl_except= (pos:822..824, aa:Thr)
XX
PN W0200005360-A1.
XX
PD 03-FEB-2000.
XX
XX 20-JUL-1999; 99WO-CA000656.
PF

XX 20-JUL-1998; 98CA-02237701.
 PR 10-DEC-1998; 98CA-02253647.
 XX (HSCR-) HSC RES & DEV LP.
 XX Wraha JL;
 XX WPI; 2000-182691/16.
 DR P-PSDB; AAY44750.
 XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 PT developing agents for treating e.g. wounds, arthritis, fibrosis,
 PT inflammation, tumors, fibrogenesis or tissue damage.
 XX Claim 8; Page 53-55; 93pp; English.
 XX The present sequence is the cDNA encoding the human Smad Anchor for
 CC Receptor Activation protein-2 (hsARA2), a TGF-beta modulator. This has
 CC been identified in the EST clone, KIAA0305 and is related to hSARA1. The
 CC hSARA proteins bind to receptor-regulated Smad proteins and ensures its
 CC appropriate localisation for activation by a Type I receptor of a TGF-
 CC beta, activin or bone morphogenetic protein (BMP) signalling pathway.
 CC SARA proteins has anti-inflammatory, cytoskeletal, anti-arthritis and
 CC vulnary activity. The SARA protein is useful for developing agents for
 CC prevention or treatment of disorders involving TGF-beta superfamily
 CC member signalling pathway. Such agents can be used for wound healing,
 CC scarring, arthritis and fibrosis (e.g., liver and kidney), in modulating
 CC inflammatory and immune responses, tumour progression, cell proliferation
 CC or fibrogenesis and in tissue morphogenesis. It is also used for
 CC protecting, restoring and regenerating tissues after tissue damage
 XX Sequence 6632 BP; 2197 A; 1123 C; 1253 G; 2059 T; 0 U; 0 Other;
 SQ

Query Match 11.6%; Score 560.6; DB 3; Length 6632;
 Best Local Similarity 60.8%; Pred. No. 5e-143;
 Matches 990; Conservative 0; Mismatches 619; Indels 18; Gaps 4;

XX 2824 CTTATCTCTGAAGTGGCTTCTCCCACTCTCATCTCCATCGTGTAAAGG---AGAC 2880
 DB 3294 CTAATGATGAGGACAGTTGCCCCCACTTCTGTTGATCTGGAGAAAGGATCAGTG 3353
 QY 2881 TATGCTGTGGAAGAAACCATCATCAGATTTTCAGTAATGCAGCAGTTGGAGGATGGTGC 2940
 DB 3354 CTTGTAGTAGAAGAACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGGTGAAGGC 3413
 QY 2941 CTGACCCACTGTATTTGTTTAAATGCAATTTGTTGCAATGTTAAATTTGTAAT 3000
 DB 3414 TTTCACTCTGTATCTTGTCTAAATGCTAAATCTACTCGTGAATGTCAAAATTCATATTT 3473
 QY 3001 TATGTGAACAGGAAGTGTGTGTGTTTCCAAACCAAGGGAATGCATGAGTGGGTCACTCT 3060
 DB 3474 TATTCCTCAGACAAATATGCTACTTTTCAACCAATGATTCATGGCTTGGGACAGGCA 3533
 QY 3061 GAGATAGTCATCTTCTACAGTGTTCACGAGTAAAGTGTTCGCAAGGATATCTTTT 3120
 DB 3534 GAAATATTATTTCTATTTGTTGTTTCCAAATGAAGATACTATTCCTAAGGACATCTTC 3593
 QY 3121 AATCACTTTTGCAGCTTTATCGGGATGCTCTGCGAGGGAATGTGGTGACCACTTGGGA 3180
 DB 3594 AGACTATTTATCACCATAATAGAGTGTCTTAAAGGAAATACATAGAAACTTTGGAC 3653
 QY 3181 CATTCCTTTCTAGTCAAGTTTCCTTGGCAGTAAGAAACATGGTGGATCTTTATATGTG 3240
 DB 3654 AATATTACCTTTTACTGAGAGTTTCTCAGTAGCAAGGATCACGAGGATTCCTGTTTAT 3713
 QY 3241 ACATCTACCTACCACTCACTGCAAGACCTAGTACTCCCAACCCCACTTACTTGTGGG 3300
 DB 3714 ACACCTACTTTTCAGAACTTGATGATCTCTCATACCAAGTAATCTTTCTTTTGGGA 3773
 QY 3301 ATTCTTATCCAGAAATGGGAAATCTTGGGCTAAAGTATTTCTATCCGCTCTGATGTTG 3360
 DB 3774 ATTCTTATCCAGAGCTTGAGATTCCTCGGCAAGGTTTTTCTATGCTTAAATGTTG 3833

QY 3361 AGACTTTGGAGCTGATATATCGACTTTTATCCATGCCCACTATTTCAGTGTGAGATTTTCGGAAG 3420
 DB 3834 AGATTGGGTGCAATATTAAGCATATCTGCTCTCTTAACAGCATCAGAGCCCAAAA 3893
 QY 3421 CCATTGTTTGGAGAGAGGGGCMATCCATCATGAATCTTCTTGCAGACTTTCAGAAATTTAC 3480
 DB 3894 CCTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACTTGTGACCTTCGAAATTTAC 3953
 QY 3481 CAGTATACCTTGGCAGTGTTCAGAGTTTGGTGGTTGATATGAAGTTTCGGAATACTAGC 3540
 DB 3954 CAGTATACCTTGGCAGTGTTCAGTGTTCAGTGTTCAGTGTTCAGTGTTCAGTGTTCAG 4013
 QY 3541 ATCAAAATTTCCAGCAACAGATATCAATGAGATGATGAAGCCATGAACAAAGTCCAATGAG 3600
 DB 4014 ATAAAAATACCAAGGAAAGTACAGTGTATTAATGMAAGTACTAAATCTTCTCAATGAG 4073
 QY 3601 CATGCTCTGCAGAGGTGCTCTTCAATGAAGGAGAGCTCTCACTTGTGTGTGA 3660
 DB 4074 CATGCTATGAGATTTGGAGCAAGTTTCAGTACAGAGGAGATTTCTATCTAGTCTGTATA 4133
 QY 3661 CAGATGATGATGGAATCTATCAGACCCAGGCTATCAGTATTTCACAATCAGCCCAAGAAA 3720
 DB 4134 CAGA---ATGATGGAATTTATGAACACAGGCGCCACAGTGGCCATCTTAGAAAA 4190
 QY 3721 GTGACTGTGCGCAGTTCTTCTGTGTTCAGTGGGCTCTGAATCTCTTCTGGATACCTT 3780
 DB 4191 GTGACAGGTGCAAGTTTGTGTGTATTCATGAGGCTCTTAAACACATCTTCAGGATTTCTT 4250
 QY 3781 GCCAAGTCCAGTATTTGTGGAAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3840
 DB 4251 GCTAAGTCCAGCATAGTTGAAGATGCTTAAATGTTAATGTTAATGTTAATGTTAATG 4310
 QY 3841 TCTTTGAGCAGGCTCTGAGAGATGAAGAGATTTCAACATCACCTGTGGGAAGCGGAC 3900
 DB 4311 GGTCTGCGCTAGCTTTACGAGACAGAGAAAGACTTTAAATTTACATGTGGGAAGTTGAT 4370
 QY 3901 GGGGAGGAACCCAGAGGACATCCACATCCAGTGGGTGGATGATGACAAGAGCTTTAGC 3960
 DB 4371 GAGTAGACCTGAGAGATACCTGGATATCTGCTGGGTAGATGCTGAAGAAAGGAAAC 4430
 QY 3961 AAGGGTGTCTAAGTCTTATAGATGGGAAGTCCATGGAGACTATTAACAAATGTGAAGATA 4020
 DB 4431 AAGGAGTTATCAGTTCAGTGTGATGATATCATTTACAAGGATTTTCAAGTGAAAAATA 4490
 QY 4021 TTCCATGAGTATGAATTAAGCAATGGAAGAAAGTAAATCAGATGAGACAGAGTGTTC 4080
 DB 4491 AAACCTGGAAGCAGATTTTGAACACCGATGAGAAGATTTGTAATAATGTACCGAGGTCTTAC 4550
 QY 4081 CTAGAAAACGATGACCAAGCAATTTGCCCTCAGTGTATCTGCAGATCAAGTAGATTGACT 4140
 DB 4551 TTTCTAA-----AGGACCAAGATTTATCTATTTTATCACTTCTTATCAGTTTGA 4601
 QY 4141 GAGCATGTTGCCAAGCTTTTGTCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCT 4200
 DB 4602 AAGAAATAGCCATGGCTTGTAGTGTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 4661
 QY 4201 GATGAAATGACCAACTGGGACTAGTGTGACACTTGTGAGGCTCAGATCAGGTTGGCTATCAA 4260
 DB 4662 ANTGGATGAATAAATTTGGACTCAGATTTCCATTTGACATGATGTTGTTGATTTGAT 4721
 QY 4261 GAGGAGCAATGCGCAGCCCTTCCCTTCGAGTACATGAATGATCTGGATAGCGCTTG 4320
 DB 4722 CGAGGATCTGAAGGCCAACTTCTGCTCTCAGCATTTATCTAAATGATCTTGTAGTGTCTG 4781
 QY 4321 GTGCGGTGTATCCATGGAGGCGCTGCGAGCTTAGTGAAGGCGCCGTTGTCTCATGAACTC 4380
 DB 4782 ATACCTGTGATCCATGGTGGAGCTTCAACTCTAGT---TTACCAATTAGAAATAGAATTA 4838
 QY 4381 ATCTTTTATTTCTGGAACACATCTATAAACAGAGAGAGCTTCAATTTTTTCTGTTTCTG 4440
 DB 4839 GTGTTTTTCAATTAGAACATCTTTTTTAGTGAAGAAATGTGCCATATTACATATTGCAA 4898

Qy	4441	ACTTGTT	4447	CC	hypergammaglobulinaemia, Crohn's disease, sarcoidosis, Gaucher's disease;
Db	4899	CCTAATT	4905	CC	cardiovascular disease e.g. cardiovascular abnormalities, persistent
				CC	truncus arteriosus, arrhythmias, angina pectoris, myocardial infarction,
				CC	hypertension, telangiectasia, ischaemia includes cerebral ischaemia,
				CC	mucocutaneous lymph node syndrome; disorders which involve angiogenesis, benign
				CC	with neovascularisation; cancers which involve angiogenesis, benign
				CC	tumours, ocular angiogenic diseases, hyperproliferative disorders,
				CC	diabetic retinopathy, uveitis, rheumatoid arthritis, psoriasis, delayed
				CC	wound healing, endometriosis, vasculogenesis, atherosclerosis; nervous
				CC	system diseases, degeneration associated with Parkinson's disease,
				CC	Alzheimer's disease and infectious diseases. This sequence encodes a
XX				CC	novel human secreted protein, described in the invention
XX				XX	
SQ	Sequence	5402 BP; 1823 A; 929 C; 1049 G; 1601 T; 0 U; 0 Other;			
	Query Match	11.6%; Score 559; DB 6; Length 5402;			
	Best Local Similarity	60.8%; Pred No. 1.2e-142;			
	Matches	989; Conservative 0; Mismatches 620; Indels 18; Gaps 4;			
Qy	2824	CTTATTCCTGAAGATGGCTTCTCCCAATTCATCTCCATCTGCTGGTGTGAAAGG----	AGAC	2880	
Db	3157	CCTAATGATGAGCAGATTTGCCCCCACTTCTGTTGCATCTGGAGAAAAGGATCAGTG		3216	
Qy	2881	TATGCTGTGGAGAGAACCACTACAGATTTTCAGTAATGCAGCAGTTGGAGATGGTGGC		2940	
Db	3217	CCTGTAGTAGAAGACATCCATCTCATGAGCAGATCATTTTCTTCTTGAAGTGAAGC		3276	
Qy	2941	CCTGACCCACATTTGTTTAAATGCAATTTTGTCAATGTTTAAATTTGTAAT		3000	
Db	3277	TTTCATCTGTTTACATTTGCTCTAAATGCTAATCTACTCTGATGATGTCATATTCATATTT		3336	
Qy	3001	TATGTGAACAGGAAGTGGTGGTGTTCACCAACAGGAATGCATGCAGTGGTGGT		3060	
Db	3337	TATTCCTCAGACAAATATTTGGTACTTTTCAACCAATGGATTCGATGGTGGGACAGGA		3396	
Qy	3061	GAGATAGTCAATCTCTACAGTGTTCACCGATGAAAGTGTTCGCCAAGGATATCTTT		3120	
Db	3397	GAAATTAATTTCTATTGTTATGTTTGGCAATGAGATACATTTCTTCTAAGGACATCTTC		3456	
Qy	3121	AATCACTTTTGCAGCTTTATCGGAGTCTCGGAGGGAATGTGGTGAAGCACTTGGGA		3180	
Db	3457	AGACTATTTTATCCCATATATATAGGATGCTCTAAAGGAAAAATACATAGAAAACCTTGGAC		3516	
Qy	3181	CATTCCTCTTCAGTCAAAAGTTTCTTGGCAGTAAGAACATGTTGATTTCTTATATGTG		3240	
Db	3517	AATATTACCTTTACTGAGAGTTTCTCAGTAGCAGGATCAGGAGGATTCCTGTTTATT		3576	
Qy	3241	ACATCTACCTACCAAGTCACTGCAAGACCTAGTACTCCCAACCCCACTTGTGTTGGG		3300	
Db	3577	ACACCTACTTTTCAGAAACTTGTATGATCTCTCATTTACCAAGTAATCCTTTTCTTTGTGA		3636	
Qy	3301	ATTCTTTATCCAGAAATGGGAACCTCTTGGGCTAAAGTATTTTCCCTATCCGTCGTGATGTG		3360	
Db	3637	ATTCTTTATCCAGAACTTGAGATTCCTCTGGGCAAGGTTTTCCTATGCTTTAATGTTG		3696	
Qy	3361	AGACTTGGAGCTGAATATCGACTTTTATCCATGCCACTATTTCAGTGTCCAGATTCGGAAG		3420	
Db	3697	AGATTGGTGCAGATATTAAGCATATCTGCTCTTAAAGCATCAGAGCCGAGAAA		3756	
Qy	3421	CCATTGTTGGAGAGACGGGGCATACCATCATGAATCTTCTTGCAGACTTCAGAAATTAC		3480	
Db	3757	CCTCTTTTGGAGAAATAGGACACACTATTATGAACCTTACTTGTGACCTTCGAAATTAC		3816	
Qy	3481	CAGTATACCTTGGCCAGTAGTTCAAGGTTTGGTGGTTGATGATGGAAGTTCGGAACATAGC		3540	
Db	3817	CAGTATACCTTGGATTAATATAGTCAACTGTTGATTCATATGAAATGGGAAAAAGCTGC		3876	
Qy	3541	ATCAAAATTTCCAGCAACAGATACAAATGAGATGATGAAAGCCATGAAAGTCCCAATGAG		3600	
Db	3877	ATAAATATACCAACGGAAAAAGTACAGTGATGTAATGAAAGTACTAAATTTCTTCCAATGAG		3936	
Qy	3601	CATGTCCTGCAGAGGTGCTCTCAATGAAAGGCGAGACTCTCATCTTGTGTGTGA		3660	

Db 3937 CATGTCATTAGCATGGAGCAAGTTTTCAGTACAGAGCAGATTCTCATCTAGTCTGTATA 3996
 QY 3661 CAGAATGATGATGGAACTATCAGACCCAGGCTATCATGATTTCAATATCAGCCAGAAAA 3720
 Db 3997 CAGA---ATGATGGAAATTTTGAACACACAGGCCAACAGCCGACTGGCCATCTTAGAAAA 4053
 QY 3721 GTGACTGGTCCAGTTTCTTTGTGTTGTCAGTGGCGCTCTGAATCTCTCTCTGATACCTT 3780
 Db 4054 GTGACAGGTGCAAGTTTGTGTTATTCATGAGCTCTAAACATCTTCAGGATTTCTT 4113
 QY 3781 GCCAGTCCAGTATGTTGAAGATGGTGTATGCTCCAGATTACTGACAGAAATGAT 3840
 Db 4114 GCTAAGTCAGCATAGTTGAAGATGGCTTAATGGTACAAATAACTCCAGAGACCATGAAT 4173
 QY 3841 TCCCTTGAGCGAGGCACTGCGAGAGATGAAGGACTTCAACATCACCCTGTGGGAAGCGGAC 3900
 Db 4174 GGCCTTGGCGTAGCTTTAGCGAACACAGAAAGACTTTAAATTTACATGTGGGAAGTTGAT 4233
 QY 3901 GCGGAGGAACCCAGGAGCAGATCCACATCCAGTGGGTGGATGATGACAAAGCTTTAGC 3960
 Db 4234 GCAGTAGACCTGAGAGAAATACGTGGATATCTGCTGGGTAGATGCTGAAGAAAAAGGAAAC 4293
 QY 3961 AAGGCTGTGTAAGTCTTATAGATGGGAAGTCCATGGAGACTATAACAAATGTGAAGATA 4020
 Db 4294 AAAGGATTTACGTTCACTGATGGATGATATCATTTACAGGATTTCCAGTGAAGAAATA 4353
 QY 4021 TTCCATGGATCAGATATAAGCAAAATGGAAGTAATCATGATGACAGAGGTGTTTTC 4080
 Db 4354 AAACCTGGAAGCAGATTTTGAACCCGATGAGAAGATTGTAAATGTACCGAGGTGTTCTAC 4413
 QY 4081 CTAGAAACGATGACACCAAGTTGCTCAGTGTCTGAGATCCGAGATCAGATGATGACT 4140
 Db 4414 TTCTTAA-----AGACACAGATTTATCTATTTTATCACTTTTACGTTTGCA 4464
 QY 4141 GAGCATGTTGCAAGCTTTTTCCTGCTCTCTGCTCCTCACTCAAACTTCTGAAGGAA 4200
 Db 4465 AAAGAAATAGCCATGGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4524
 QY 4201 GATGAATGACCAACTGGGACTAGCTGTGACACTTGTGATCAGATCAGTGGTGGTATCAA 4260
 Db 4525 AATGGGATGAATAAATTTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 4584
 QY 4261 GCAGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGATGATGATGATGATGATGATG 4320
 Db 4585 GCAGGATCGAAGCCACTTCTGCTCAGATTTATTAATGATCTGATGATGATGATGATG 4644
 QY 4321 GTGCGGTGATCCATGGAGGGCTGCGAGCTTGTAGTGGGCCCTGTTGTCATGGAATCT 4380
 Db 4645 ATACCTGTGATCCATGGTGGGACCTCCAACTCTAGT---TTACCAATTAGAAATAGAA 4701
 QY 4381 ATCTTTTATATCTGGAACACATGTTATTAACAGAGAGACTTCATTTTTCGTTTCAG 4440
 Db 4702 GTGTTTTCATATAGAACATCTTTTATGTAAGAAATGTCGATATATACATATTCGAA 4761
 QY 4441 ACTTGTT 4447
 Db 4762 CCTAATT 4768

RESULT 14

AAZ50071
 ID AAZ50071 standard; cDNA; 2678 BP.

XX AC AAZ50071;

XX DT 04-MAY-2000 (first entry)

XX DE Xenopus Smad Anchor for Receptor Activation protein-2 encoding cDNA.

XX KW Smad Anchor for Receptor Activation protein; XSARA2; TGF-beta;
 transforming growth factor-beta; bone morphogenetic protein; BMP;
 activin; anti-inflammatory; cytosolic; antiarthritis; vulnary;
 KW TGF-beta modulator; wound healing; scarring; arthritis; immune response;

KW inflammatory response; tumour progression; cell proliferation; fibrosis;
 fibrogenesis; tissue morphogenesis; tissue damage; ss.

XX Xenopus sp.

XX FH Key Location/Qualifiers

XX CDS 347..2032

XX FT /tag= a

XX FT /product= "Xenopus SARA-2 protein"

XX FT /note= "Binds to receptor regulated Smad proteins"

XX PN W0200005360-A1.

XX PD 03-FEB-2000.

XX PF 20-JUL-1999; 99NO-CA000656.

XX PR 20-JUL-1998; 98CA-02237701.

XX PR 10-DEC-1998; 98CA-02253647.

XX PA (HSCR-) HSC RES & DEV LP.

XX PI Wraha JL;

XX DR WPI; 2000-182691/16.
 P-PSDB; AAY44752.

XX New Smad Anchor for Receptor Activation (SARA) polypeptides, useful for
 developing agents for treating e.g. wounds, arthritis, fibrosis,
 inflammation, tumors, fibrogenesis or tissue damage.

XX PS Claim 10; Page 60; 93pp; English.

XX The present sequence is the cDNA encoding the Xenopus Smad Anchor for
 Receptor Activation protein-2 (XSARA2), a TGF-beta modulator. The XSARA
 proteins bind to receptor-regulated Smad proteins and ensures its
 appropriate localisation for activation by a Type I receptor of a TGF-
 beta, activin or bone morphogenetic protein (BMP) signalling pathway.
 SARA proteins has anti-inflammatory, cytostatic, anti-arthritis and
 vulnary activity. The SARA proteins are useful for developing agents
 for prevention or treatment of disorders involving TGF-beta superfamily
 member signalling pathway. Such agents can be used for wound healing,
 scarring, arthritis and fibrosis (e.g. liver and kidney), in modulating
 inflammatory and immune responses, tumour progression, cell proliferation
 or fibrogenesis and in tissue morphogenesis. It is also used for
 protecting, restoring and regenerating tissues after tissue damage

XX Sequence 2678 BP; 875 A; 482 C; 532 G; 789 T; 0 U; 0 Other;

Query Match 6.9%; Score 333.2; DB 3; Length 2678;

Best Local Similarity 64.6%; Pred No. 1.6e-80;

Matches 567; Conservative 0; Mismatches 293; Indels 18; Gaps 4;

QY 1443 ACAGTGTGTGGATGGCAGATGCGAGTCTAGATTAAAGAACTTCGATAGGAAG 1502

Db 1111 AGAGTTGGAAGGAGGGTAGATGCTAATCTTATGGAAGATGATGATGATGATGATG 1170

QY 1503 TGAAGATGCTGATTTCTCCACTGTTATAGACACACCCAGCAGCAATATCTATCTATGG 1562

Db 1171 AGAAGAATAGATTGTTCTGGAAGAAATGGAACCTAATGCAATGTTATCTGATCATGG 1230

QY 1563 TTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCAGAGACTTTTACCCTC 1622

Db 1231 GTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTGTTCAGAGACTTTTACCCTC 1290

QY 1623 CAAAGATTCAGTAAACAGAGAAATAGAGAGAAATAGAGAGAAATAGAGAGAAATAGAG 1682

Db 1291 AAAAGAGATTCCTGACAGAGAAATAGAGAGAAATAGAGAGAAATAGAGAGAAATAGAG 1350

QY 1683 AAATATTTATG---AACAGAGAGAAATAGAGAGAAATAGAGAGAAATAGAGAGAAATAGAG 1739

Db 1351 TGGTCTTTATGAACACAGAGAAATAGAGAGAAATAGAGAGAAATAGAGAGAAATAGAG 1410

QY 1740 CAGCACTGGTGAAGCAAAATATTTTACATATTTCTGTAGTCAAGTTCCATC 1799
Db 1411 TGGCAAGGTTGACCAAAATGAAGACA---GTTTGCATAGTCTTTAATCCGGTTCATC 1467
QY 1800 AGTGTCTGGGCAATCTTCCGCCAGGTAGTAGCAAGCTGCCATCTATCAGTGTTCCTTT 1859
Db 1468 CATGTCATGGGCAACCTCACCAAAAAGGCAAGATTGTGCAATCCCTCAGTGTTCATA 1527
QY 1860 TGGTGTGCAAGACCAAGCAACTTCTTAATCTTAACCTCAATTTCCAAAGCCATTATC 1919
Db 1528 TGGTGGAGCTCGGCCAAGCAGCAACTCTCAACTCAATTTCCACAGCCATTGTC 1587
QY 1920 AGACCAATTTCAAAATCA---CTTCTCTGCAAAACAGTGGAAATAATACTAAAAATAAAA 1976
Db 1588 TGAATGTTACAGTGTGATCTCATTCGCCCAATGCTGGATGAGCTCTTAATAACAAAA 1647
QY 1977 TGATATTCTTTGGGAAACAAAATTTAGGGGAAACTCAGCAACCAATGTATGCAAGTCCATC 2036
Db 1648 TGACATGTTAAACAAATCAAAATCGGGGGGATAACCTGATTTCAAGATCACTACGTGAGGA 1707
QY 2037 TTTGGGAAACATCTCTAATCTCGATACAAATGGGACATTTTAGAAAGTTATGAGGCTGA 2096
Db 1708 AGTCACAGCCCTGTTACTGATACAAATGGTGAAGTCCCTCGAGAAACAGGGGACCTGG 1767
QY 2097 GATCTCCACTAGACCATGCTTGCATTAGCTCCAGATAGCCAGATATGATCTCAGAGC 2156
Db 1768 CAGC-----CTGTGCTTGCAGTGTCTCCAGACAGCCCTGACAAATGATCTGCTTGC 1818
QY 2157 TGGTCAAGTTTGGAAATTTTGGCCAGAAAGCCATTCACACGCTGGTGGAGTCCAGT 2216
Db 1819 TGGACAGTTTGGGGTACCCATCTCTAAGCCATTTTACTACTAGGGGATGTGGCTCCAGT 1878
QY 2217 ATGGGTACCGGATTTCTCAGGCTCCAAATTTGATGAAATGGAAGCCAGGTTTACATTAC 2276
Db 1879 CTGGGTGCCAGATTCCCAAGCACCACCACTGATGATGAGTGGAGGCCAGATTACATTAC 1938
QY 2277 CAAAAGGAGGCATCACTGCAAGAGCATGTGGGAAGGTTT 2314
Db 1939 CAAAAGGAGGCATCACTGCGGAGCTGTGGAAAGGTAT 1976

RESULT 15
AAC30390 standard; cDNA; 332 BP.
AAC30390;
06-OCT-2000 (first entry)
Human secreted protein 5' EST, SEQ ID NO: 34465.
Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
gene therapy; chromosome mapping; ss.
Homo sapiens.
EP1033401-A2.
06-SEP-2000.
21-FEB-2000; 2000EP-00200610.
26-FEB-1999; 99US-0122487P.
(GEST) GENSET.
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 34465; 71pp + Sequence Listing; English.
PS The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. NO ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX
SQ Sequence 332 BP; 105 A; 67 C; 70 G; 90 T; 0 U; 0 Other;
Query Match 6.6%; Score 318; DB 3; Length 332;
Best Local Similarity 99.7%; Pred. No. 7.5e-77;
Matches 329; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 171 CAAGGTGTAGATGAATTTGAACAAACGAGATCAACAGTTCTTCTTACTTTATTGA 230
Db 1 CAAGGTGTAGATGAATTTGAACAAACGAGATCAACAGTTCTTCTTACTTTATTGA 60
QY 231 TACAAAGTGAATAGATTTCTAGATCCCTCTCTCACCGGCTGTCAATTTAACTTCTT 290
Db 61 TACAAAGTGAATAGATTTCTAGATCCCTCTCTCACCGGCTGTCAATTTAACTTCTT 120
QY 291 GGCC-AGTGTGAATGAATCTGCAGTTTCTAATGAGTCAACACCACTGAAAGTCTTCT 349
Db 121 GGCCAAAGTGAATGAATCTGCAGTTTCTAATGAGTCAACACCACTGAAAGTCTTCT 180
QY 350 CCTGGCTCATTTGAGTCCCTGACACAGAGCAAGAGGATCACTGTGTAATGGACAGG 409
Db 181 CCTGGCTCATTTGAGTCCCTGACACAGAGCAAGAGGATCACTGTGTAATGGACAGG 240
QY 410 ACTGTAATCTAAATCCAGAGATTGCCCAATGTGAATGATGAAATGCTGTTCGAGAAG 469
Db 241 ACTGTAATCTAAATCCAGAGATTGCCCAATGTGAATGATGAAATGCTGTTCGAGAAG 300
QY 470 ACCAGTTAATTAAGAGAACTATAGTTGGG 499
Db 301 ACCAGTTAATTAAGAGAACTATAGTTGGG 330

Search completed: May 17, 2004, 18:15:06
Job time : 1124.1 secs

This Page Blank (uspto)

375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Forward

POLYA-yes

Location/Qualifiers

1. 706

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-EJO-ahk-b-16-0-UI"

/tissue_type="fetal eyes, lens, eye anterior segment,

optic nerve, retina, Retina Foveal and Macular, RPE and

Choroid"

/dev stage="fetal and adult"

/lab host="DH10B (Life Technologies) (T1 phage resistant)"

/clone lib="UI-E-EJO"

/note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-EJO is a subtracted cDNA library constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pTV73-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tags for this library are: fetal eyes,

AGAATCAGCA; lens, CGATTAGCGA; eye anterior segment,

AATGCCGAT; optic nerve, CCAATTAAGTG; retina, CCGCG; Retina

Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This

library was created for the program, Gene discovery in the

Visual System, supported by National Eye Institute (NEI).

TAG_TISSUE=Foveal and Macular Retina

TAG_LIB=UI-E-EJO

TAG_SEQ=GTCC"

ORIGIN

Query Match 14.4%; Score 699.2; DB 12; Length 706;
Best Local Similarity 99.4%; Pred. No. 9.1e-128;
Matches 701; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 950 AGGGACAGATGGATGCTCTGTGTTAAAAAGCAAGAGAACTATATACCATGAGGACC 1009

DB 706 AGGGACAGATGGATGCTCTGTGTTAAAAAGCAAGAGAACTATATACCATGAGGACC 647

QY 1010 TCACTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTCCACA 1069

DB 646 TCACTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTCCTTTCCACA 587

QY 1070 TGAGTGGGGATTTTGTATGAAAAAGACGACAGAGAGAGACCACTGAAGAATCCC 1129

DB 586 TGAGTGGAGNATTTTGTATGAAAAAGACGACAGAGAGAGACCACTGAAGAATCCC 527

QY 1130 TCCGCTCTGGTTTACCTTTGCTTCTCAAAACGACATCCCTAATGGGTCTGGAGGAATA 1189

DB 526 TCCGCTCTGGTTTACCTTTGCTTCTCAAAACGACATCCCTAATGGGTCTGGAGGAATA 467

QY 1190 ATGACTGTGAACGGTTCAGATTCCTGTGCTTAATCAAGTTAGGCTCATGAATG 1249

DB 466 ATGACTGTGAACGGTTCAGATTCCTGTGCTTAATCAAGTTAGGCTCATGAATG 407

QY 1250 AAGGTTATCAACATCAAGAAACTCTTGGCACTACAGAAATTCCTTAATATGACAGCAT 1309

DB 406 AAGGTTATCAACATCAAGAAACTCTTGGCACTACAGAAATTCCTTAATATGACAGCAT 347

QY 1310 TCTCTGAATCTCAGGACATGACTAATTTGGAAGTTGACTAAATAATGAGATGATGATA 1369
DB 346 TCTCTGAATCTCAGGACATGACTAATTTGGAAGTTGACTAATGAGATGATGATA 287
QY 1370 GCCAAGTAAACCAAGAAAGGAAAAAGTTTCTACAGATTAGTCAGCTCAGGACACTAATG 1429
DB 286 GCCAAGTAAACCAAGAAAGGAAAAAGTTTCTACAGATTAGTCAGCTCAGGACACTAATG 227
QY 1430 GTGATAGTGGAGGACAGAGTGTGTGGATTGGCAGATGCGAGTCTAGATTAAAGGAAT 1489
DB 226 GTGATAGTGGAGGACAGAGTGTGTGGATTGGCAGATGCGAGTCTAGATTAAAGGAAT 167
QY 1490 GCATTAGTGAAGAGTGAAGAAATGTTCTCCACTGTTTATAGACACACAGCAGCAAAAT 1549
DB 166 GCATTAGTGAAGAGTGAAGAAATGTTCTCCACTGTTTATAGACACACAGCAGCAAAAT 107
QY 1550 ATCTATCTAATCGTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTTCCTCA 1609
DB 106 ATCTATCTAATCGTTGTGATTCCTATGGAATGCAAGACCCAGGTGTTCTTTTTCCTCA 47
QY 1610 AGACTTTACCTCCCAAGAGAGATTTCAGTAACAGAGAAAAAGAAA 1654
DB 46 AGACTTTACCTCCCAAGAGAGATTTCAGTAACAGAGAAAAAGAAA 2

RESULT 6

LOCUS BQ421818

DEFINITION BQ421818

ACCESSION BQ421818

VERSION BQ421818.1

KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 941)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLAM13218 row: o column: 13
High-quality sequence stop: 609.

FEATURES

Source

1. 941

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6018276"

/tissue_type="epithelioid carcinoma"

/lab host="DH10B (phage-resistant)"

/clone lib="NIH MGC 70"

/note="Organ: pncrfas; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.1 kb. Library constructed by Life

Technologies."

ORIGIN

Query Match 14.4%; Score 699.2; DB 13; Length 941;

Best Local Similarity 98.0%; Pred. No. 8.8e-128;

Matches 729; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 3433 GAGACGGGGCATACCATCATCAATCTCTTTCGACAGATTCAGAAATACCATGATACCTTG 3492


```

Qy 4761 TGGGTACCAACCATCACCAAGGGTGGGATGGAGGGGAGGGGAGGGAATATAAAGCA 4820
Db 73 TGGGTACCAACCATCACCAAGGGTGGGATGGAGGGGAGGGGAGGGAATATAAAGCA 14
Qy 4821 TCAAAAAAAAAA 4833
Db 13 TCAAAAAAAAAA 1

RESULT 8
LOCUS BI334801 751 bp mRNA linear EST 30-JUL-2001
DEFINITION BI334801 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141187 5',
mRNA sequence.
ACCESSION BI334801
VERSION BI334801.1 GI:15019458
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM11347 row: n column: 04
High quality sequence stop: 742.
FEATURES
    source
        1..751
            Location/Qualifiers
                1..751
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5141187"
                    /tissue_type="cervical carcinoma cell line"
                    /lab_host="DH10B"
                    /clone_lib="NIH_MGC_12"
                    /note="Organ: Cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                    Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                    Average insert size 1.4 kb. Library prepared by Life
                    Technologies."
ORIGIN
Query Match 13.4%; Score 648.6; DB 12; Length 751;
Best Local Similarity 94.3%; Pred. No. 9.1e-118;
Matches 708; Conservative 0; Mismatches 34; Indels 9; Gaps 3;

Qy 3238 GTGACATCTACTACAGTCACTCAAGACCTAGTACTCCCAACCCACCTTACTTGT 3297
Db 1 GTGACATCTACTACAGTCACTCAAGACCTAGTACTCCCAACCCACCTTACTTGT 60
Qy 3298 GGGATCTTATCCAGAAATGGGAACCTCTTGGGCTAAAGTATTCTATCCGTCGATG 3357
Db 61 GGGATCTTATCCAGAAATGGGAACCTCTTGGGCTAAAGTATTCTATCCGTCGATG 120
Qy 3358 TTGAGACTTGGAGCTGAATATCGATTTCATCCCACTTATTCAGTGTGAGATTTCGG 3417
Db 121 TTGAGACTTGGAGCTGAATATCGATTTCATCCCACTTATTCAGTGTGAGATTTCGG 180
Qy 3418 AAGCCATTGTTGGAGACGGGGGCATACCATCATGAATCTCTTTCGAGACTTCAGAAAT 3477
Db 181 AAGCCATTGTTGGAGACGGGGGCATACCATCATGAATCTCTTTCGAGACTTCAGAAAT 240
Qy 3478 TACCAAGTATACCTCCAGTAGTTCAAGGTTGGTGGTGTATGATGGAAGTTCGGAAC 3537

```

```

Db 241 TACCAAGTATACCTCCAGTAGTTCAAGGTTGGTGGTGTATGATGGAAGTTCGGAAC 300
Qy 3538 AGCATCAAAATTCACGCAACAGATACAATGAGATGATGAAGCCATGAACAGTCCAAAT 3597
Db 301 AGCATCAAAATTCACGCAACAGATACAATGAGATGATGAAGCCATGAACAGTCCAAAT 360
Qy 3598 GAGCATGCTCTGGCAGGAGGTCCTTCAATGAAAAGGAGGAGCTCTCATCTTGTGTGT 3657
Db 361 GAGCATGCTCTGGCAGGAGGTCCTTCAATGAAAAGGAGGAGCTCTCATCTTGTGTGT 420
Qy 3658 GTACAGATGATGATGGAACCTATCAGACCCAGGCTATCAGTATTCAAAATCAGCCAGA 3717
Db 421 GTACAGATGATGATGGAACCTATCAGACCCAGGCTATCAGTATTCAAAATCAGCCAGA 480
Qy 3718 AAAGTGACTGTGTCAGATTCTTT--GTGTTTCAGTGGGCTCTG-AAATCCTCTCTCGA 3774
Db 481 AAAGTGACTGTGTCAGATTCTTTCTCGTGTTCAGTGGGCTCTGCAAAATCTCTTCTTGA 540
Qy 3775 TACCTTGCCAACTCCAGTATTGTG-----GAAGATGGTGTATTGTCGCAGATTACTGCA 3828
Db 541 TACCTTGCCAACTCCAGTATTGTGTTGGAACGATGCTGTCTATCGTCCACATTACTGCA 600
Qy 3829 GAGAACTGATGATTCCTTGGAGCAGGACTGCGAGAGATGAAGACTTCACCATCACCCTGT 3888
Db 601 GAGAACTGATGATTCCTTGGAGCAGGACTGCGAGAGATGAAGACTTCACCATCACCCTGT 660
Qy 3889 GGGAGCGCGAGCGGAGGAAACCCAGGAGCACATCCACATCCAGTGGGTGATGATGAC 3948
Db 661 GGGAGCGCGAGCGGAGGAAACCCAGGAGCACATCCACATCCAGTGGGTGATGATGAC 720
Qy 3949 AAGAACTTTAGCAAGGGTGTCTGAAGTCTTA 3979
Db 721 AAGAACTTTAGCAAGGGTGTCTGAAGTCTTA 751

RESULT 9
LOCUS BM716723 669 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-B-EJO-abk-b-16-0-UI.12 UI-E-EJO Homo sapiens cDNA clone
ACCESSION BM716723 1 GI:19029981
VERSION BM716723.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 669)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse
Location/Qualifiers
    1..669
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
FEATURES
    source
        1..669
            Location/Qualifiers

```



```

VERSION BX096354.1 GI:27842763
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 644)
JOURNAL Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
COMMENT Radlof, U., Schneider, D. and Korn, B.
Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGP998L141206.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl.cgi/response?libNo=972
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
Contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTACACAGAAACAGCATGAC.
Location/Qualifiers
1..644
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGP998L141206 ; IMAGE:505501"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NbHPU"
/site="Organ: uterus; Vector: pT7T3-Pac; Site:1: Not I;
Site:2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer 15',
AACTGGAGAATTCGCGCCGCTTTTTTTTTTTT 3' ],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
FEATURES
source
Query Match 13.0%; Score 631.4; DB 13; Length 644;
Best Local Similarity 99.1%; Pred. No. 2.4e-114;
Matches 632; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 940 ATCTCATCCAGGACAGATGGATGCTCTGCTGTAAAGACGAGAGCAATATACCA 999
DB 7 ATCTCATCCAGGACAGATGGATGCTCTGCTGTAAAGACGAGAGCAATATACCA 66
QY 1000 GATGAGGACCTCAGCTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTC 1059
DB 67 GATGAGGACCTCAGCTGGCAAAATCAGCTCTCTAGGACAGATCTAGGAGTCCAAATTC 126
QY 1060 TTTTCCACATGAGTGAGGGGATTTGATGAAAAAGACGAGAGGAGGAGCACCCT 1119
DB 127 TTTTCCACATGAGTGAGGGGATTTGATGAAAAAGACGAGAGGAGGAGCACCCT 186
QY 1120 GAAGAATCCCTCCGCTCTGTTTACCTTTGCTTCTCAACAGACATGCTTAATGGTCT 1179
DB 187 GAAGATCCCTCCGCTCTGTTTACCTTTGCTTCTCAACAGACATGCTTAATGGTCT 246
QY 1180 GGAAGGAATATGACTGGAACGGTGTTCAGATTGCCTTGTGCTTAAGTTAGGGCT 1239
DB 247 GGAAGGAATATGACTGGAACGGTGTTCAGATTGCCTTGTGCTTAAGTTAGGGCT 306
QY 1240 GATGAAATGAAGTTATGACATGAGAACTCTTGGCACTACAGATTCCTTAATG 1299

```

us-09-744-167-1.rst

Page 10

Db 307 GATGAAATGAAGTTATGACATGAGAACTCTTGGCACTACAGATTCCTTAATG 366

QY 1300 ACAGAGATTTCCTGAACTCAGGACATGAACTTGAAGTTGAACTAAATGAG 1359

Db 367 ACAGAGATTTCCTGAACTCAGGACATGAACTTGAAGTTGAACTAAATGAG 426

QY 1360 ATGAATCATAGCAAGTAAACGAGAAAGGAAAGTTCTACAGATTACTGAGCTGAG 1419

Db 427 ATGAATCATAGCAAGTAAACGAGAAAGGAAAGTTCTACAGATTACTGAGCTGAG 486

QY 1420 GACACTAATGTTGATAGTGGAGGACAGATGTTGGATTGGCAGATCGAGTCTAGATT 1479

Db 487 GACACTAATGTTGATAGTGGAGGACAGATGTTGGATTGGCAGATCGAGTCTAGATT 546

QY 1480 AAGGAACTTGCATTAGTGAAGTGAAGATGATGTTCTCCACTGTTATAGACACCA 1539

Db 547 AAGGAACTTGCATTAGTGAAGTGAAGATGATGTTCTCCACTGTTATAGACACCA 606

QY 1540 GCAGCAAAATATCTATCTAATGTTGATTCCTATGG 1577

Db 607 GCAGCAAAATATCTATCTAATGTTGATTCCTATGG 644

RESULT 13

CB049545

LOCUS NISC_gj11e08.y1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3271406

DEFINITION 5', mRNA sequence.

ACCESSION CB049545

VERSION CB049545.1 GI:277787832

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 654)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Plate: LLAN8008 row: J column: 15
Seq primer: M13Rpi reverse primer (ABI).
Location/Qualifiers
1..654
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3271406"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr28"
/note="Organ: prostate; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Pr22 was prepared, and ss circles were used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

```

http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyX-5,
Location/Qualifiers
1. 777
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6819584"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_EXO"
/note="Organ: Brain; Vector: pyX-Asc; Site_1: EcoR I;
Site_2: Not I; the library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:7791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pyX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGCAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

```

3598	QY	GAGCGGAGGAACTCCAGGAGCAATCCACATCCAGTGGGAGATGATGACAAAGACGTT	3599
181	Db	GATGCGAGAGACCCGAGGAACAGATCCATCCAGTGGGTGGACGATGACAGACAGTT	240
3958	QY	AGCAAGGGTGTCTTAAGTCTCTATAGATGGGAAGTCCATGGAGACTATAAACAAATGTGAAG	4017
241	Db	AACAAAGGTGTGTGAGTCTCTATAGATGGGAAGTCTATGAATCTATAACAAAGTGAAA	300
4018	QY	ATATTCATGGATCAGAAATAAAGCAAAATGGAAAAAGTAAATCAGATGGGACAGAGTGTTT	4077
301	Db	ATATTCATGGGTGAGATATAAGCAAAACGAAAAAGTCAATCAGATGGACAGAGGTGTTT	360
4078	QY	TTCTGAAAAACGATGACCAGCAAAATTCGCTTCAGTGTATCCTGCAGATCACAGTATGATG	4137
361	Db	TTCTGTGAAACGATGACCACTAATCTGCTGAGTGACCTTCGAGATCACAGCAGACTG	420
4138	QY	ACTGAGCATGTTGCCAAAGCTTTTTCGCTTGCTCTCTGTCTCTCACTGAAACTTCTGAAG	4197
421	Db	ACTGAGCATGTGCGCAAGGCATTCTGCTCGCCCTCTGCCCCACCTTGAAGCTTTTAAAA	480
4198	QY	GAGATGGAATGACCAACTGGGACTACGTGTGACACTTCGACTCAGATCAGGTGGGTAT	4257
481	Db	GAGATGGAATGACAAACTGGGACTCCGTGTGACCGCTTGACCTCAGATCAGGTGGGTAT	540
4258	QY	CAAGCAGGGAGCAATGGCCAGCCCTTCCCTCGCAGTACATGAATGATCTCGATAGGCC	4317

Db 541 CAAGCAGGAGCAATGGCCAGCCCTTCCCTGCGAGTACATGACGATCTGACAGTGCC 600
 QY 4318 TTGGTGCCGGTGATCCATGAGGGGCGCTGCGAGCTTAGTGTAGGGCCCGCTGTGATGAA 4377
 Db 601 TTGGTGCCGTGCTCAATTCATGAGGAGCGCTGCCAGCTCAGTGAAGGCCCTGTGTCATGGA 660
 QY 4378 CTCACCTTTTATATCTGGAACAACTCGTATAAACAGAGAAGACTTCATTTTCTCTT 4437
 Db 661 CTCATCTTTTATCTGGAACAACTCGCAT----AGATGAGGACTCCATTTGTTCTGTT 716
 QY 4438 CAGACTTGTGCAACAGCAGTCAATCCCAATTCATTTGACCTTTAAACTGGA 4490
 Db 717 NCAGACTGTTGAGCAGCAGTCAATCCANATCAATTCGCAATANACTGAA 769

RESULT 15
 BU745833
 LOCUS
 DEFINITION
 Canis familiaris cDNA clone CH2#002_G03 5', mRNA sequence.
 ACCESSION
 BU745833
 VERSION
 EST..
 SOURCE
 Canis familiaris (dog)
 ORGANISM
 Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 806)
 Y. Y., Desai, R., Olarte, M., Henthorn, P. and George A. L.
 Expressed sequence tags from Canine heart
 Unpublished (2003)
 Other ESTs: CH2#002_G03T3
 Contact: George A. L.
 Division of Genetic Medicine
 Vanderbilt University
 529 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0275, USA
 Tel: 615 936 2660
 Fax: 615 936 2661
 Email: al.george@vanderbilt.edu
 Insert Length: 1688 Std Error: 0.00
 Seq primer: T7: TAATACGACTCACTATAGG
 High quality sequence start: 51
 High quality sequence stop: 573.
 Location/Qualifiers
 1..806
 /organism="Canis familiaris"
 /mol_type="mRNA"
 /db_xref="taxon:9615"
 /clone="CH2#002_G03"
 /tissue_type="heart"
 /cell_type="heart"
 /dev_stage="mixed developmental stages (adult, 30 day - 40 day fetal)"
 /clone_lib="Canine heart normalized cDNA Library in pBluescript"

FEATURES
source

Query Match 12.6%; Score 609.4; DB 13; Length 806;
 Best Local Similarity 88.8%; Pred. No. 5.1e-110;
 Matches 682; Conservative 0; Mismatches 82; Indels 4; Gaps 2;
 QY 1859 TTGGTGGTGCAAGACCCAGCAACCTTCTAACTTTAAACTTCAAATTCAAAGCCATTAT 1918
 Db 34 TTGNGGTCTCGAACCAGCAACCTTCTAACTTTAAACTTCAAATTCATAGCCATTGT 93
 QY 1919 CAGACCAATTTCAAAATGACTTCTTCGCAACAGTGGAAATTAATAAATAAATG 1978

ORIGIN

Query Match 12.6%; Score 609.4; DB 13; Length 806;
 Best Local Similarity 88.8%; Pred. No. 5.1e-110;
 Matches 682; Conservative 0; Mismatches 82; Indels 4; Gaps 2;
 QY 1859 TTGGTGGTGCAAGACCCAGCAACCTTCTAACTTTAAACTTCAAATTCAAAGCCATTAT 1918
 Db 34 TTGNGGTCTCGAACCAGCAACCTTCTAACTTTAAACTTCAAATTCATAGCCATTGT 93
 QY 1919 CAGACCAATTTCAAAATGACTTCTTCGCAACAGTGGAAATTAATAAATAAATG 1978

Db 94 CTGACCATTTGCAAAATGACCTTCTTACAAATCGTGGAAATTAATTTAAATAAATAAATG 153
 QY 1979 ATATTTCTTGGGAAGCAAAATAGGGGAACCTCAGCAACCAATGTATCAGCTCCATCTT 2038
 Db 154 ATGTTCTTGGGAAGCAAAATAGGGGAACCTCAGCAACCAATGTATGCAATGCTCTT 213
 QY 2039 TGGGAACATCTCTAATGTGATACAAATGGGAACATTTAGAAAGTTATAGAGGTGAGA 2098
 Db 214 TGGGAACATGTCGGTTGCTGATACAAATGTTTAAACATTTAGATACTACGAGTCTGGGA 273
 QY 2099 TCTCCACTAGACCATGCTTGCATAGCTCCAGATAGCCAGATATGATCTCAGAGCTG 2158
 Db 274 TATCCAGTAGACCGTGCCTTGCATAGCTCCAGAAAGCCAGATATGATCTCAGAGCTG 333
 QY 2159 GTCAGTTTGAATTTCTGCCAGAAAGCCATTCACACGCTGGGTGAGGTGCCAGTAT 2218
 Db 334 GTCATTTGGAATTTCTGCCAGAAAGCCATTTACTACTCTGGGTGAGGTGCCAGTAT 393
 QY 2219 GGATACCGGATTTCTAGGCTCCAAATTTGCATGAATGTGAAGCCAGGTTTACATTCACCA 2278
 Db 394 GGGTACCAATTTCTAGGCTCCAAATTTGCATGAATGTGAAGCCAGGTTTACATTCACCA 453
 QY 2279 AAGGAGGATCACTGCAGAGCATGTGGGAAGGTTTTCTGCTCTCTCTGTAGCTCA 2338
 Db 454 AAGGAGGATCACTGCAGAGCATGTGGGAAGGTTTTCTGCTCTCTCTGTAGCTCA 513
 QY 2339 AATGTAACCTGTTTATACATGGACAGAAAGAGCTAGAGTGTGTGTAATCTGCCATTCAG 2398
 Db 514 AATGTAACCTGTTTATACATGGATAGAAAGAGGCTAGAGTGTGTGTAATCTGCCATTCAG 573
 QY 2399 TGCTAATGATGCTCAAGCTTGGGAGAAATGATGATGCTCAAGCCAGAGCCCTAACCC 2458
 Db 574 TGCTAATGATGCTCAAGCTTGGGAGAAATGATGATGCTCAAGCCAGAGCCCTAACCC 633
 QY 2459 CTAAACATCTGCTGAATACTGTTCTACTATCCCTTTCAGCAAGCTCAGGCTCTCA 2518
 Db 634 CTAAACATCTGCTGAATACTGTTCTACTATCCCTTTCAGCAAGCTCAGGCTCTCA 693
 QY 2519 GAGCTGTGAGCTCTCCACCTCCCACTGTGATGCTGAGGAGTTTAAAGCACCCTG 2578
 Db 694 GAAACCTGAGCTCTCCACCTCCCACTGTGATGCTGAGGAGTTTAAAGCACCCTG 753
 QY 2579 GAG---CAGAGTGGCTCAGCCC-AGAGAGCAGAGCGAGTTTGGTTT 2622
 Db 754 TGGAGCAAGATGTGGCTCAACCAAGAGCAGAGCGGATTTTGGTT 801

Search completed: May 18, 2004, 07:03:20
 Job time : 7424.14 secs